Statistica Sinica **24** (2014), 447-461
doi:http://dx.doi.org/10.5705/ss.2011.129

A SIMPLIFICATION OF COMPUTING MARKOV BASES FOR GRAPHICAL MODELS WHOSE UNDERLYING GRAPHS ARE SUSPENSIONS OF GRAPHS

Shoufeng Cai[1,2], Benchong Li[1] and Jianhua Guo[1]

[1]*Northeast Normal University and* [2]*Jilin University*

Abstract: In this paper, we extend the framework of Dobra and Sullivant (2004) for computing Markov bases for some models, compute Markov bases for model selections by applying our method, and analyse a contingency table.

Key words and phrases: Graphical model, Markov basis, saturation of ideal, suspension of graph.

1. Introduction and Notation

The concept of a Markov basis comes from hypothesis tests for contingency tables in statistics (Diaconis and Sturmfels (1998)). A Markov basis can be applied to Fisher's exact test for some contingency tables in which the Chi-square test is deficient, such as contingency tables containing cells with frequencies less than 5 (Haberman (1988)), or, the one in Wermuth and Cox (1998). The fundamental theorem of Markov bases (Diaconis and Sturmfels (1998)) states that computation of a Markov basis for a log-linear model is equivalent to computing a Groebner basis (Cox, Little, and O'Shea (2007)) for a toric ideal describing the log-linear model. The techniques used to compute Markov bases involve computational algebraic geometry tools. Drton and Sullivant (2007) formulated the relations between affine varieties and exponential family, and gave some examples to illustrate applications to some problems in statistical inference for algebraic models. Fienberg (2007) summarized the development of algebraic statistics. Drton, Sturmfels, and Sullivant (2009) formulated some basic ideas in algebraic statistics and gave directions for future research. Riccomagno (2009) presented a brief history of algebraic statistics.

De Loera and Onn (2006) showed that computation of a Markov basis for a log-linear model is an NP-hard problem. However, Markov bases for some special models can be computed, especially for some undirected graphical models (Lauritzen (1996)). In this paper, we consider undirected graphs.

First we review some concepts for graphs. The vertex set of a graph $\mathbf{G}$ is denoted by $\mathcal{V}(\mathbf{G})$. If $V_1, V_2 \in \mathcal{V}(\mathbf{G})$ and are connected in $\mathbf{G}$, we denote the edge that connects them by the unordered pair (V_1, V_2) and the edge set of $\mathbf{G}$ by $\mathcal{E}(\mathbf{G})$. Given $V \in \mathcal{V}(\mathbf{G})$, we define $ne(V) = \{W \in \mathcal{V}(\mathbf{G}) \mid (V, W) \in \mathcal{E}(\mathbf{G})\}$ (Lauritzen (1996)). A vertex $V \in \mathcal{V}(\mathbf{G})$ is called a universal vertex of $\mathbf{G}$ if $\mathcal{V}(\mathbf{G}) = \{V\} \cup ne(V)$.

If a graph $\mathbf{H}$ can be decomposed as $\mathbf{H}_1$ and $\mathbf{H}_2$ (Tarjan (1985); Leimer (1993); Lauritzen (2002)), we write $\mathbf{H} = \mathbf{H}_1 \oplus \mathbf{H}_2$ and call $\mathbf{H}$ reducible. If the graph is not reducible, we call it a prime graph. Given two graphs $\mathbf{G}, \mathbf{H}$, if $\mathcal{V}(\mathbf{H}) \subseteq \mathcal{V}(\mathbf{G}), \mathcal{E}(\mathbf{H}) \subseteq \mathcal{E}(\mathbf{G})$, we call $\mathbf{H}$ a subgraph of $\mathbf{G}$.

Next, we define the operation on two graphs denoted by $\otimes$.

Definition 1. If graphs $\mathbf{H}_1, \mathbf{H}_2$ have $\mathcal{V}(\mathbf{H}_1) \cap \mathcal{V}(\mathbf{H}_2) = \emptyset$, the graph $\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2$ satisfies $\mathcal{V}(\mathbf{H}) = \mathcal{V}(\mathbf{H}_1) \cup \mathcal{V}(\mathbf{H}_2), \mathcal{E}(\mathbf{H}) = \mathcal{E}(\mathbf{H}_1) \cup \mathcal{E}(\mathbf{H}_2) \cup \{(V_1, V_2) \mid V_1 \in \mathcal{V}(\mathbf{H}_1), V_2 \in \mathcal{V}(\mathbf{H}_2)\}$. If $\mathcal{V}(\mathbf{H}_1) = \{V\}, \mathbf{H}_1 \otimes \mathbf{H}_2$ is called a suspension of $\mathbf{H}_2$ over V (Sturmfels and Sullivant (2008)) and V is a universal vertex of $\mathbf{H}_1 \otimes \mathbf{H}_2$.

Graphs that contain a universal vertex have many applications, such as latent class models (Fienberg et al. (2010)) and toric geometry of cuts and splits (Sturmfels and Sullivant (2008)).

Example 1. We can see that E is a universal vertex of $\mathbf{G}_1$ (Figure 1), A is a universal vertex of $\mathbf{G}_2$ (Figure 2), and the vertex E is universal for the graph G_3 (Figure 1.3).

Dobra (2003) constructed a Markov basis for a graphical model corresponding to a graph $\mathbf{G}$ that is decomposable (Lauritzen (1996)). For reducible $\mathbf{G}$, Dobra and Sullivant (2004) proved that a Markov basis for the model can be constructed from Markov bases for smaller models whose underlying graphs are prime subgraphs $\mathbf{G}_1, \ldots, \mathbf{G}_n$ of graph $\mathbf{G}$. Hence, we need only compute Markov bases for graphical models whose underlying graphs are prime. Develin and Sullivant (2003) described a Markov basis for a binary graphical model whose underlying graph is an N-cycle or a complete bipartite graph. Král', Norine, and Pangrác (2010) studied the width of the Markov basis for some graphical models. The website `http://markov-bases.de/` has a database of Markov bases for some log-linear models.

In this paper, we present a method for simplifying the computation of Markov bases for a class of graphical models. If $\mathbf{G}$ is a prime graph that contains a universal vertex V and $\mathbf{G} = \mathbf{H} \otimes \{V\}$, we can reduce the problem of computing a Markov basis for the graphical model of $\mathbf{G}$ to computation of a Markov basis for the graphical model of $\mathbf{H}$. Since the number of vertices is smaller for $\mathbf{H}$ than for $\mathbf{G}$, the computational complexity is reduced. Based on this method, we obtain an extension of the framework of Dobra and Sullivant (2004) for computing Markov

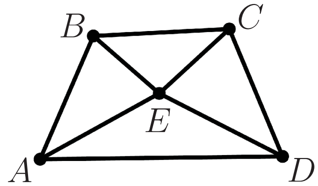


Figure 1. Graph $\mathbf{G}_1, ne(E)=\{A,B,C,D\}$.

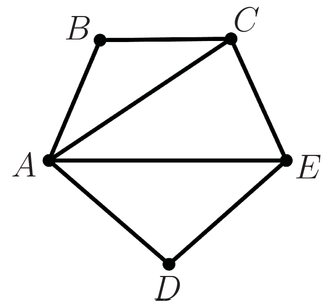


Figure 2. Graph $\mathbf{G}_2, ne(A)=\{B,C,D,E\}$.

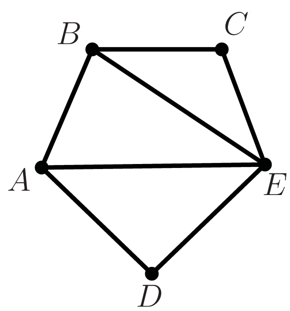


Figure 3. Graph $\mathbf{G}_3, ne(E)=\{A,B,C,D\}$.

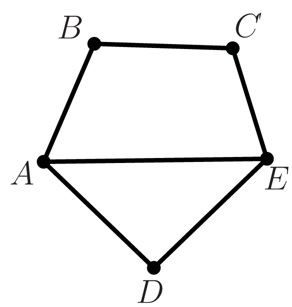


Figure 4. Graph $\mathbf{G}_4, ne(E)=\{A,C,D\}$.

Table 1. Contingency table.

A.B.	C.D.E							
	111	112	121	122	211	212	221	222
11	928	699	1132	422	53	26	85	26
12	232	130	295	77	8	3	7	0
21	180	388	592	432	12	12	38	16
22	26	60	81	74	0	1	4	0

bases for some models; this extension widens the application areas for Markov bases.

We embed our method in the method of Krampe and Kuhnt (2010) to analyse a contingency table. We select a graphical model with the best fit to the contingency data among graphical models for graphs $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ and $\mathbf{G}_4$ (Figures 1, 2, 3 and 4). We use our method to compute a Markov basis for the graphical model for $\mathbf{G}_1$. The $4 \times 5 \times 5 \times 2 \times 2$ contingency table provided by Wermuth and Cox (1998) is converted to a $2 \times 2 \times 2 \times 2 \times 2$ contingency table (Table 1) by level aggregation, using the method of Dellaportas and Tarantola (2005). We discuss this problem in Example 5.

The remainder of the paper is organized as follows. In the next section, we provide a method to simplify the computation of a Markov basis for a graphical model whose underlying graph contains a universal vertex. In Section 3, the relationship between reducibility and a universal vertex for a graph is analysed and three examples are presented. Section 4 concludes. The appendix lists computational algebraic geometry tools that are used to prove our main result.

Let $\mathbb{N}$ be the set of positive integers, $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. If $m, n \in \mathbb{N}$, then let $[n] = \{1, \ldots, n\}$ and let $\mathbf{K}_m$ denote the complete graph with m vertices. Given an index set U, we take $X_U = (x_\alpha)_{\alpha \in U}$, in this paper, the index set is the cells of a contingency table (Geiger, Meek, and Sturmfels (2006)).

2. Markov Basis for Graphical Models Corresponding to $\mathbf{G} \otimes \mathbf{K}_m$

This section describes the simplification method used to reduce the computational basis for a toric ideal corresponding to a graphical model whose underling graph contains a universal vertex. The properties of a graphical model whose underlying graph $\mathbf{G}$ has a universal vertex help in constructing a basis for the toric ideal $I_{\mathbf{G}}$ according to the graphical model. According to the fundamental theorem for Markov bases (Diaconis and Sturmfels (1998)), we need only compute the basis for the corresponding toric basis. Since the saturation of a homogeneous binomial ideal is still a homogeneous binomial ideal (Eisenbud and Sturmfels (1996), every toric ideal $I_{\mathbf{G}}$ has a basis that consists of homogeneous binomials.

Given a contingency table, Hoşten and Sullivant (2002) introduced the tableau notion for monomials to formulate the relations between the cells of the contingency table more clearly. They associate each monomial $x_{\alpha_1} \cdots x_{\alpha_t}$, where $\alpha_j = (a_{j1}, \ldots, a_{jn})$ is the index of a cell in the contingency table for $j = 1, \ldots, t$, t is the total degree of the monomial, n is the number of random variables, with a $t \times n$ tableau

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{t1} & \ldots & a_{tn} \end{pmatrix}.$$

If a variable occurs to its $m-$th power in a monomial, its corresponding index set occurs m times in the tableau.

Example 2. For a binary graphical model whose underlying graph is a four-cycle graph $\mathbf{G}$, where $\mathcal{V}(\mathbf{G}) = \{A, B, C, D\}$, the relations of conditional independence among A, B, C, D can be formulated by binomials, and one of the binomials

$$x_{1212}x_{1111} - x_{1112}x_{1211}$$

can be rewritten in tableau notion as

$$\begin{pmatrix} 1\,2\,1\,2 \\ 1\,1\,1\,1 \end{pmatrix} - \begin{pmatrix} 1\,1\,1\,2 \\ 1\,2\,1\,1 \end{pmatrix}.$$

Theorem 1. *For a graphical model whose underlying graph $\hat{\mathbf{G}}$ has vertex set $\{V_1, V_2, \ldots, V_{n+1}\}$, let $d_i \in [\gamma_i]$ be the level value of the random variable V_i, where $\gamma_i \in \mathbb{N}, i \in [n+1]$. Suppose V_{n+1} is a universal vertex in $\hat{\mathbf{G}}, \hat{\mathbf{G}} = \mathbf{G} \otimes \{V_{n+1}\}$ and $\{f_1, f_2, \ldots, f_l\} \subseteq \mathbb{R}[X]$ is a basis for the toric ideal $I_{\mathbf{G}}$, where*

$$f_i = \begin{pmatrix} a_{i,1,1} & \ldots & a_{i,1,n} \\ \vdots & \ddots & \vdots \\ a_{i,k,1} & \ldots & a_{i,k,n} \end{pmatrix} - \begin{pmatrix} b_{i,1,1} & \ldots & b_{i,1,n} \\ \vdots & \ddots & \vdots \\ b_{i,k,1} & \ldots & b_{i,k,n} \end{pmatrix},$$
$$i \in [l], X = (x_\alpha), \alpha \in \mathcal{S} = [\gamma_1] \times \cdots \times [\gamma_n].$$

Then $\bigcup_{e=1}^{\gamma_{n+1}} \{f_1^{(e)}, f_2^{(e)}, \ldots, f_l^{(e)}\} \subseteq \mathbb{R}[\tilde{X}]$ is a basis for the toric ideal $I_{\hat{\mathbf{G}}}$, where

$$f_i^{(e)} = \begin{pmatrix} a_{i,1,1} & \ldots & a_{i,1,n} & e \\ \vdots & \ddots & \vdots & \vdots \\ a_{i,k,1} & \ldots & a_{i,k,n} & e \end{pmatrix} - \begin{pmatrix} b_{i,1,1} & \ldots & b_{i,1,n} & e \\ \vdots & \ddots & \vdots & \vdots \\ b_{i,k,1} & \ldots & b_{i,k,n} & e \end{pmatrix},$$
$$i \in [l], e \in [\gamma_{n+1}], \tilde{X} = (\tilde{x}_\beta), \beta \in \mathcal{T} = [\gamma_1] \times \cdots \times [\gamma_{n+1}].$$

The proof is in the Appendix.

From Theorem 1, computation of a Markov basis for a graphical model with underlying graph $\hat{\mathbf{G}} = \mathbf{G} \otimes \{V\}$ can be reduced to a simpler model whose underlying graph is $\mathbf{G}$. If the number of universal vertices of $\hat{\mathbf{G}}$ is greater than 1, then we can generalize Theorem 1.

Lemma 1. *Let $\mathbf{G}$ be a graph with $\mathcal{V}(\mathbf{K_{m+1}}) \cap \mathcal{V}(\mathbf{G}) = \emptyset, V \in \mathcal{V}(\mathbf{K}_{m+1})$. If $\mathbf{K}_{m+1} = \mathbf{K}_m \otimes \{V\}$, then $\mathbf{G} \otimes \mathbf{K}_{m+1} = (\mathbf{G} \otimes \mathbf{K}_m) \otimes \{V\}$.*

Proof. The result is obvious from the definition of $\otimes$.

Based on Theorem 1 and Lemma 1, we have the following.

Corollary 1. *Suppose we have a graphical model whose underlying graph $\hat{\mathbf{G}}$ has vertex set $\{V_1, V_2, \ldots, V_{n+m}\}$. Let $\mathbf{K}_m$ be the complete subgraph that satisfies $\hat{\mathbf{G}} = \mathbf{G} \otimes \mathbf{K}_m, \mathcal{V}(\mathbf{K}_m) = \{V_{n+1}, \ldots, V_{n+m}\}$, where $d_i \in [\gamma_i]$ are the level values of V_i, $i \in [n+m]$. Suppose $\{f_1, f_2, \ldots, f_l\} \subseteq \mathbb{R}[X]$ is a basis for the toric ideal $I_{\mathbf{G}}$, where*

$$f_i = \begin{pmatrix} a_{i,1,1} & \ldots & a_{i,1,n} \\ \vdots & \ddots & \vdots \\ a_{i,k,1} & \ldots & a_{i,k,n} \end{pmatrix} - \begin{pmatrix} b_{i,1,1} & \ldots & b_{i,1,n} \\ \vdots & \ddots & \vdots \\ b_{i,k,1} & \ldots & b_{i,k,n} \end{pmatrix},$$

$$i \in [l], , X = (x_\alpha), \alpha \in \mathcal{S} = [\gamma_1] \times \cdots \times [\gamma_n].$$

Then

$$\bigcup_{\epsilon \in [\gamma_{n+1}] \times \cdots \times [\gamma_{n+m}]} \{f_1^{(\epsilon)}, f_2^{(\epsilon)}, \ldots, f_l^{(\epsilon)}\} \subseteq \mathbb{R}[\tilde{X}]$$

is a basis for the toric ideal $I_{\hat{\mathbf{G}}}$, where

$$f_i^{(\epsilon)} = \begin{pmatrix} a_{i,1,1} & \ldots & a_{i,1,n} & \epsilon \\ \vdots & \ddots & \vdots & \vdots \\ a_{i,k,1} & \ldots & a_{i,k,n} & \epsilon \end{pmatrix} - \begin{pmatrix} b_{i,1,1} & \ldots & b_{i,1,n} & \epsilon \\ \vdots & \ddots & \vdots & \vdots \\ b_{i,k,1} & \ldots & b_{i,k,n} & \epsilon \end{pmatrix},$$

$$i \in [l], \epsilon \in [\gamma_{n+1}] \times \cdots \times [\gamma_{n+m}], \tilde{X} = (\tilde{x}_\beta), \beta \in \mathcal{T} = [\gamma_1] \times \cdots \times [\gamma_{n+m}].$$

3. Simplification of a Graphical Model Class

We present some graph theory results, use the results of Dobra and Sullivant (2004) to discuss the problem of computing Markov bases for some special graphical models. Three examples are given.

Proposition 1. *Let* $\mathbf{H}$ *be a graph. If* V *is a universal vertex of* $\mathbf{H}$ *and* $\mathbf{H} = \mathbf{H}_1 \oplus \mathbf{H}_2$*, then* V *is a universal vertex of both* $\mathbf{H}_1$ *and* $\mathbf{H}_2$*.*

Proof. Since V is a universal vertex of $\mathbf{H}$, then V is connected to all other vertices of $\mathbf{H}$ and thus $V \in \mathcal{V}(\mathbf{H}_1) \cap \mathcal{V}(\mathbf{H}_2)$. Then V is a universal vertex of both $\mathbf{H}_1$ and $\mathbf{H}_2$.

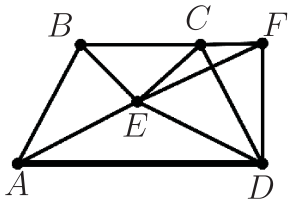
Proposition 2. *Let* $\mathbf{H}$ *be a graph. If* $\hat{\mathbf{H}} = \mathbf{H} \otimes \{V\}$*, then* $\hat{\mathbf{H}}$ *is reducible if and only if* $\mathbf{H}$ *is reducible.*

Proof. We can verify the equivalence between $\mathbf{H} = \mathbf{H_1} \oplus \mathbf{H_2}$ and $\hat{\mathbf{H}} = \hat{\mathbf{H}}_\mathbf{1} \oplus \hat{\mathbf{H}}_\mathbf{2}$ by reducibility.

In fact, Propositions 1 and 2 are valid for finite universal vertices.

Now, combining the results of Dobra and Sullivant (2004), we extend the framework to reduce the complexity of computing a basis for the toric ideal $I_\mathbf{G}$. If a graph $\mathbf{G}$ is reducible, it is first decomposed into its prime subgraphs. If any of the prime subgraphs has a universal vertex, we simplify this subgraph and the process continues until a prime graph without a universal vertex is obtained. Then the problem of computing a basis for the toric ideal $I_\mathbf{G}$ is converted to small toric ideals corresponding to prime subgraphs of $\mathbf{G}$.

Example 3. Consider a graphical model with binary-valued nodes for graph $\mathbf{G_5}$ (Figure 5). By the discussion, graph $\mathbf{G_5}$ can be decomposed to graphs $\mathbf{G_1}$ (Figure 1) and $\mathbf{G_6}$ (Figure 6); since $\mathbf{G_1}$ has a universal vertex E, $\mathbf{G_1}$ can be simplified



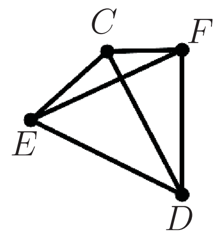


Figure 5. Graph $\mathbf{G}_5, ne(E)=\{A,B,C,D,F\}$. Figure6. Graph $\mathbf{G}_6, ne(E)=\{C,D,F\}$.

to a four-cycle prime graph $\mathbf{G}$, where $\mathcal{V}(\mathbf{G}) = \{A, B, C, D\}$. By Theorem 1 the problem of computing a Markov basis for graphical model according to G_1 can be simplified to $\mathbf{G}$. Then a Markov basis for the graphical model of $\mathbf{G_5}$ can be constructed using the method of Dobra and Sullivant (2004), in addition to that $\mathbf{G_6}$ is a complete graph.

Example 4. Let $\mathbf{G}$ be a four-cycle graph, $\mathcal{V}(\mathbf{G}) = \{A, B, C, D\}, \mathbf{G}_1 = \mathbf{G} \otimes \{E\}$ (Figure 1). By considering binary graphical models (Geiger, Meek, and Sturmfels (2006)), a basis can be obtained for the toric ideal $I_{\mathbf{G}}$:

$$
\begin{aligned}
&x_{2122}x_{2221} - x_{2121}x_{2222}, x_{1222}x_{2212} - x_{1212}x_{2222}, x_{2112}x_{2211} - x_{2111}x_{2212},\\
&x_{1221}x_{2211} - x_{1211}x_{2221}, x_{1122}x_{2112} - x_{1112}x_{2122}, x_{1122}x_{1221} - x_{1121}x_{1222},\\
&x_{1112}x_{1211} - x_{1111}x_{1212}, x_{1121}x_{2111} - x_{1111}x_{2121},\\
&x_{1211}x_{1222}x_{2112}x_{2121} - x_{1212}x_{1221}x_{2111}x_{2122},\\
&x_{1121}x_{1212}x_{2122}x_{2211} - x_{1122}x_{1211}x_{2121}x_{2212},\\
&x_{1112}x_{1221}x_{2121}x_{2212} - x_{1121}x_{1212}x_{2112}x_{2221},\\
&x_{1112}x_{1222}x_{2121}x_{2211} - x_{1122}x_{1212}x_{2111}x_{2221},\\
&x_{1111}x_{1122}x_{2212}x_{2221} - x_{1112}x_{1121}x_{2211}x_{2222},\\
&x_{1111}x_{1222}x_{2112}x_{2221} - x_{1112}x_{1221}x_{2111}x_{2222},\\
&x_{1111}x_{1222}x_{2122}x_{2211} - x_{1122}x_{1211}x_{2111}x_{2222},\\
&x_{1111}x_{1221}x_{2122}x_{2212} - x_{1121}x_{1211}x_{2112}x_{2222}.
\end{aligned}
$$

Using Theorem 1, we can obtain a basis for the toric ideal $I_{\mathbf{G}_1}$ that consists of 16 quadratic binomials and 16 quartic binomials, provided in http://markov-bases.de/ (SPg_bin). The basis is

$$
\begin{aligned}
&x_{21221}x_{22211} - x_{21211}x_{22221}, x_{12221}x_{22121} - x_{12121}x_{22221},\\
&x_{21121}x_{22111} - x_{21111}x_{22121}, x_{12211}x_{22111} - x_{12111}x_{22211},\\
&x_{11221}x_{21121} - x_{11121}x_{21221}, x_{11221}x_{12211} - x_{11211}x_{12221},
\end{aligned}
$$

Table 2. Levels for the variables A–E.

A	B	C	D	E
$i=1$, very poorly	$j=1$, basic incomplete	$k=1,\ 19-29$	$l=1,\ 1991$	$r=1$, West
$i=2$, poorly	$j=2$, basic	$k=2,\ 30-44$	$l=2,\ 1992$	$r=2$, East
$i=3$, well	$j=3$, medium	$k=3,\ 45-59$		Germany
$i=4$, very well	$j=4$, upper medium	$k=4,\ 60-74$		
	$j=5$, intensive	$k=5,\ \geq 75$		

$$
\begin{aligned}
&x_{11121}x_{12111}-x_{11111}x_{12121}, x_{11211}x_{21111}-x_{11111}x_{21211},\\
&x_{21222}x_{22212}-x_{21212}x_{22222}, x_{12222}x_{22122}-x_{12122}x_{22222},\\
&x_{21122}x_{22112}-x_{21112}x_{22122}, x_{12212}x_{22112}-x_{12112}x_{22212},\\
&x_{11222}x_{21122}-x_{11122}x_{21222}, x_{11222}x_{12212}-x_{11212}x_{12222},\\
&x_{11122}x_{12112}-x_{11112}x_{12122}, x_{11212}x_{21112}-x_{11112}x_{21212},\\
&x_{12111}x_{12221}x_{21121}x_{21211}-x_{12121}x_{12211}x_{21111}x_{21221},\\
&x_{11211}x_{12121}x_{21221}x_{22111}-x_{11221}x_{12111}x_{21211}x_{22121},\\
&x_{11121}x_{12211}x_{21211}x_{22121}-x_{11211}x_{12121}x_{21121}x_{22211},\\
&x_{11121}x_{12221}x_{21211}x_{22111}-x_{11221}x_{12121}x_{21111}x_{22211},\\
&x_{11111}x_{11221}x_{22121}x_{22211}-x_{11121}x_{11211}x_{22111}x_{22221},\\
&x_{11111}x_{12221}x_{21121}x_{22211}-x_{11121}x_{12211}x_{21111}x_{22221},\\
&x_{11111}x_{12221}x_{21221}x_{22111}-x_{11221}x_{12111}x_{21111}x_{22221},\\
&x_{11111}x_{12211}x_{21221}x_{22121}-x_{11211}x_{12111}x_{21121}x_{22221},\\
&x_{12112}x_{12222}x_{21122}x_{21212}-x_{12122}x_{12212}x_{21112}x_{21222},\\
&x_{11212}x_{12122}x_{21222}x_{22112}-x_{11222}x_{12112}x_{21212}x_{22122},\\
&x_{11122}x_{12212}x_{21212}x_{22122}-x_{11212}x_{12122}x_{21122}x_{22212},\\
&x_{11122}x_{12222}x_{21212}x_{22112}-x_{11222}x_{12122}x_{21112}x_{22212},\\
&x_{11112}x_{11222}x_{22122}x_{22212}-x_{11122}x_{11212}x_{22112}x_{22222},\\
&x_{11112}x_{12222}x_{21122}x_{22212}-x_{11122}x_{12212}x_{21112}x_{22222},\\
&x_{11112}x_{12222}x_{21222}x_{22112}-x_{11222}x_{12112}x_{21112}x_{22222},\\
&x_{11112}x_{12212}x_{21222}x_{22122}-x_{11212}x_{12112}x_{21122}x_{22222}.
\end{aligned}
$$

Using the result in Example 4, we analyse the contingency Table 1.

Example 5. The $4\times5\times5\times2\times2$ table is from Wermuth and Cox (1998), who investigated the factors that influence political attitudes. Response variable A is political attitude, B is the type of formal schooling, C is the age group, D is time, and E is region. The levels of the five variable are given in Table 2. We obtain the contingency table (Table 1) by level aggregation using the method of

Table 3. MLE for the graphical model of $\mathbf{G}_1$.

A.B.	C.D.E							
	111	112	121	122	211	212	221	222
11	923.76	701.58	1138.60	425.67	55.13	25.86	80.47	19.90
12	235.74	129.21	290.58	78.39	6.37	1.36	9.30	1.04
21	181.16	383.37	588.44	430.39	10.81	14.13	41.59	20.12
22	25.35	62.85	82.33	70.55	0.69	0.66	2.64	0.94

Table 4. MLE for the graphical model of $\mathbf{G}_2$.

A.B.	C.D.E							
	111	112	121	122	211	212	221	222
11	936.68	669.41	1165.29	409.61	62.28	31.17	77.48	19.07
12	216.14	154.46	268.89	94.51	5.90	2.95	7.34	1.81
21	178.38	388.58	585.05	439.99	11.86	12.78	38.89	14.47
22	27.00	58.82	88.57	66.61	0.76	0.82	2.49	0.93

Table 5. MLE for the graphical model of $\mathbf{G}_3$.

A.B.	C.D.E							
	111	112	121	122	211	212	221	222
11	918.50	698.91	1142.67	427.66	60.97	28.81	75.86	17.63
12	234.50	128.78	291.73	78.79	7.03	1.51	8.74	0.92
21	180.11	381.95	590.72	432.49	11.96	15.74	39.21	17.83
22	25.18	62.58	82.59	70.86	0.75	0.73	2.48	0.83

Dellaportas and Tarantola (2005, Sec. 5.4). The principle is to determine which graph ($\mathbf{G}_1$, Figure 1; $\mathbf{G}_2$, Figure 2; $\mathbf{G}_3$, Figure 3; $\mathbf{G}_4$, Figure 4) provides the best fit to the data.

We use the model selection approach of Krampe and Kuhnt (2010). Given a model for selection, a p-value is computed using the Metropolis–Hastings algorithm. For graphical models of $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ and $\mathbf{G}_4$, we first compute the maximum likelihood estimate using the IPS algorithm (Tables 3$-$6).

A Markov basis for the graphical model of graph G_1 is shown in Example 4. Since graphs G_2 and G_3 are decomposable and G_4 is reducible, Markov bases for graphical models of G_2, G_3 and G_4 are easy to compute. Then we obtain p-values using the Metropolis–Hastings algorithm (Table 7).

Remark 1. We compare our result in Example 5 to the conclusion drawn by (Dellaportas and Tarantola, 2005, Sec. 5.4).

We compute the p-values for the four models as

$$0.050(\mathbf{G}_2) < 0.071(\mathbf{G}_4) < 0.616(\mathbf{G}_3) < 0.696(\mathbf{G}_1).$$

Graph $\mathbf{G}_2$ strictly contains $\mathbf{G}_4$ and the p-value for $\mathbf{G}_2$ is strictly less than that for $\mathbf{G}_4$. This confirms that the model selection approach of Krampe and Kuhnt

Table 6. MLE for the graphical model of $\mathbf{G}_4$.

A.B.	C.D.E							
	111	112	121	122	211	212	221	222
11	936.91	672.93	1165.58	411.76	62.24	27.39	77.43	16.76
12	215.87	155.04	268.55	94.87	5.98	2.63	7.44	1.61
21	178.34	385.76	584.92	436.80	11.85	15.70	38.85	17.78
22	27.07	58.54	88.77	66.29	0.75	0.99	2.46	1.13

Table 7. p-values for graphical models of $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ and $\mathbf{G}_4$.

Model	$\mathbf{G}_1$	$\mathbf{G}_2$	$\mathbf{G}_3$	$\mathbf{G}_4$
p-value	0.696	0.050	0.616	0.071

(2010) is valid. The p-values for graphical models of $\mathbf{G}_1$ and $\mathbf{G}_3$ are almost the same, but $\mathbf{G}_3$ has fewer edges than $\mathbf{G}_1$ has. The graphical model for $\mathbf{G}_3$ is then selected. By utilizing the Akaike information criterion and the reversible jump algorithm on the aggregated table, $\mathbf{G}_3$ (Figure 3) is selected (Dellaportas and Tarantola (2005)), it is the same graph as our example.

In Example 5, we show that our method can be applied to model selection by using the method of Krampe and Kuhnt (2010). We use the aggregated binary data set from Wermuth and Cox (1998) that is used in Dellaportas and Tarantola (2005) for two reasons: our method is valid by using the data set; the computational complexity is low and the conclusion is easy to show. Of course, the method of Krampe and Kuhnt (2010) can be applied to the non-binary original data set in Wermuth and Cox (1998), but the result is too long to fit here, so we use the aggregated table instead.

4. Conclusion

It is hard to compute a Markov basis for a general undirected graphical model. This paper presents a simplification method for computing a Markov basis for a class of graphical models whose underlying graphs contain universal vertices. If the underlying graph can be reduced to a series of prime graphs (Dobra and Sullivant (2004)) and certain prime graphs contain universal vertices, then our method can be applied to reduce the graphical models of the prime graphs to simpler models.

Acknowledgement

The authors wish to thank the Editor, an associate editor and two referees for very helpful comments which greatly improved the presentation of this paper. This research was partially supported by the National Natural Science Foundation of China (Grant 11025102), PCSIRT, the Jilin Project (20100401), and the Science Foundation for Young Teachers of NENU (No. 20070102).

Appendix: Proof of Theorem 1

To prove Theorem 1, we first describe a condition for computing a reduced Gröbner basis for an ideal sum. Second, we prove that the computing order for two operations for an ideal sum and an ideal saturation (Cox, Little, and O'Shea (2007, p.197)) can be changed under given conditions. Finally, we prove Theorem 1.

Given a positive integer n and polynomial ring $\mathbb{R}[X], X = (x_i), i \in [n]$, we take $S, T \subseteq [n], S \neq \emptyset, T \neq \emptyset, S \cap T = \emptyset$.

Geiger, Meek, and Sturmfels (2006) presented a method for computing a basis for the toric ideal I_G from $I_{pairwise(G)}$ by ideal saturation given a graph G, that is $I_G = I_{pairwise(G)} : X^\infty$. If we know the reduced Gröbner bases of two ideals $I, J \subset \mathbb{R}[X]$ under a given condition, we can directly construct the reduced Gröbner basis for the ideal sum from the reduced Gröbner bases for I, J.

Proposition A.1. Let I and J be ideals in $\mathbb{R}[X]$. Given a monomial order $\prec$, if I has the reduced Gröbner basis $\mathcal{F} = \{f_i \in \mathbb{R}[X_S] \,|\, i = 1, 2, \ldots, l_1\}$, J has the reduced Gröbner basis $\mathcal{G} = \{g_j \in \mathbb{R}[X_T] \,|\, j = 1, 2, \ldots, l_2\}$, then the reduced Gröbner basis for $I + J$ is $\hat{\mathcal{G}} = \mathcal{F} \cup \mathcal{G}$.

Proof. Since $\mathcal{F}$ and $\mathcal{G}$ are reduced Gröbner bases, according to the definition of a reduced Gröbner basis we need only prove that $\hat{\mathcal{G}}$ is a Gröbner basis for $I + J$. We test this according to Buchberger's criterion. If f is a polynomial, we denote the leading term of f by $\mathrm{LT}(f)$.

(i) If $f_i, f_j \in \mathcal{F}$, because $\mathcal{F}$ is the reduced Gröbner basis for I and $S \cap T = \emptyset$, by the division algorithm for polynomials, we have $\overline{S(f_i, f_j)}^{\hat{\mathcal{G}}} = 0$.

(ii) The case $g_i, g_j \in \mathcal{G}$ is the same as (i).

(iii) If $f_i \in \mathcal{F}, g_j \in \mathcal{G}$, we compute $\overline{S(f_i, g_j)}^{\hat{\mathcal{G}}}$ according to the division algorithm in Cox, Little, and O'Shea (2007, Chap. 2). Without loss of generality, take $i = j = 1$. Let X^γ be the least common multiple of $\mathrm{LT}(f_1)$ and $\mathrm{LT}(g_1)$,

$$S(f_1, g_1) = \frac{X^\gamma}{\mathrm{LT}(f_1)} \cdot f_1 - \frac{X^\gamma}{\mathrm{LT}(g_1)} \cdot g_1.$$

Since $\mathcal{F}$ is a reduced basis, there are no $\mathrm{LT}(f_i), i = 2, 3, \ldots, l_1$, that divide any term in $S(f_1, g_1)$. The same is true for $\mathrm{LT}(g_i), i = 2, 3, \ldots, l_2$. Then $\overline{S(f_1, g_1)}^{\hat{\mathcal{G}}} = 0$ and thus the proposition is proved.

Corollary A.1. If I is a homogeneous ideal in $\mathbb{R}[X]$, then $I : x_1$ is homogeneous too.

Proof. Fix the grevlex monomial order induced by $x_n > \cdots > x_1$. According to Cox, Little, and O'Shea (2007, p.380, Thm. 2), I has a reduced Gröbner basis consisting of homogeneous polynomials; according to Sturmfels (1996, Lemma 12.1), $I : x_1$ is a homogeneous ideal.

We discuss the problem of how to change the computing order for an ideal sum and an ideal saturation in detail.

Proposition A.2. Let I and J be homogeneous ideals in $\mathbb{R}[X]$. Suppose that I has a basis $\mathcal{F} = \{f_i \in \mathbb{R}[X_S] \,|\, i = 1, 2, \ldots, l_1\}$ and J has a basis $\mathcal{G} = \{g_j \in \mathbb{R}[X_T] \,|\, j = 1, 2, \ldots, l_2\}$. If $1 \in S$, then $(I + J) : x_1^\infty = (I : x_1^\infty + J)$.

Proof. Fix the grevlex monomial order induced by $x_n > \cdots > x_1$. By Corollary A.1, $I : x_1$ is a homogeneous ideal. From the algorithm for computing a reduced Gröbner basis, I has a reduced Gröbner basis $\hat{\mathcal{F}} = \{\hat{f}_i \in \mathbb{R}[X_S] \,|\, i = 1, 2, \ldots, m_1\}$ and J has a reduced Gröbner basis $\hat{\mathcal{G}} = \{\hat{g}_j \in \mathbb{R}[X_T] \,|\, j = 1, 2, \ldots, m_2\}$. By Proposition A.1 the reduced Gröbner basis for $I + J$ is $\hat{\mathcal{F}} \cup \hat{\mathcal{G}}$. Using Lemma 12.1 of Sturmfels (1996), the Gröbner basis for $(I + J) : x_1^\infty$ is obtained by dividing each element $f \in \hat{\mathcal{F}} \cup \hat{\mathcal{G}}$ by the highest power of x_1 that divides f. Since $1 \notin T$, no leading term of $g \in \hat{\mathcal{G}}$ can be divided by x_1. Then $(I + J) : x_1^\infty = (I : x_1^\infty + J)$.

Proposition A.2 can be generalized from x_1 to $x_1 \cdots x_n$.

Proposition A.3. Let $I, J \subseteq \mathbb{R}[X]$ be two ideals. Suppose I has a basis $\mathcal{F} = \{f_i \in \mathbb{R}[X_S] \,|\, i = 1, 2, \ldots, l_1\}$ and J has a basis $\mathcal{G} = \{g_j \in \mathbb{R}[X_T] \,|\, j = 1, 2, \ldots, l_2\}$. Then $(I + J) : (x_1 \cdots x_n)^\infty = I : (\prod_{s \in S} x_s)^\infty + J : (\prod_{t \in T} x_t)^\infty$.

Proof. Based on the discussion following Lemma 12.1 of Sturmfels (1996),

$$(I + J) \,:\, (x_1 x_2 \ldots x_n)^\infty \;=\; ((\cdots(((I + J) \,:\, x_1^\infty) \,:\, x_2^\infty) \cdots) \,:\, x_n^\infty).$$

We compute $(I + J) : x_1^\infty$.

(i) If $1 \in S$, then $1 \notin T$ and $(I + J) : x_1^\infty = (I : x_1^\infty) + J$.

(ii) If $1 \in T$, similar to the case above, $(I + J) : x_1^\infty = I + (J : x_1^\infty)$.

(iii) If $1 \notin S \cup T$, then $(I + J) : x_1^\infty = I + J$ holds according to Sturmfels (1996, Lemma 12.1).

If we consider $x_2, \ldots, x_n$ one by one as above, the conclusion is confirmed.

Corollary A.2. Let $I_i =< f_t^{(i)} \in \mathbb{R}[X_{S_i}] \,|\, t \in [l_i] >\subset \mathbb{R}[X], i \in [k]$ be ideals, where $S_i \subset [n], S_i \neq \emptyset, S_{i_1} \cap S_{i_2} = \emptyset, i_1 \neq i_2$. Then

$$(\sum_{i=1}^{k} I_i) : (x_1 \cdots x_n)^\infty = \sum_{i=1}^{k} (I_i : (\prod_{s \in S_i} x_s)^\infty).$$

Lemma A.1 presents an algorithm for computing a basis for an ideal that satisfies special conditions. Before computation of an ideal saturation, we define a monomial order.

Definition A.1. Let $\mathbb{R}[x_1, \ldots, x_n]$ be a polynomial ring and let $\mathbb{R}[x_1, \ldots, x_m](m < n)$ be a subring of $\mathbb{R}[x_1, \ldots, x_n]$. Fix a monomial order $\prec_n$ for $\mathbb{R}[x_1, \ldots, x_n]$. A monomial order $\prec_m$ for $\mathbb{R}[x_1, \ldots, x_m]$ is $X^\alpha \prec_m X^\beta$ if $X^\alpha \prec_n X^\beta$. We call $\prec_m$ a monomial order derived from $\prec_n$.

Lemma A.1. Consider a polynomial ring $\mathbb{R}[x_1, \ldots, x_n]$ and its subring $\mathbb{R}[x_1, \ldots, x_m](m < n)$. If $I =< f_1, \ldots, f_s >$ is an homogeneous ideal in $\mathbb{R}[x_1, \ldots, x_m]$, $I' =< f_1, \ldots, f_s >$ is an homogeneous ideal in $\mathbb{R}[x_1, \ldots, x_n]$. Given a monomial order $\prec_n$ for $\mathbb{R}[x_1, \ldots, x_n]$, $\prec_m$ is a monomial order derived from $\prec_n$ for $\mathbb{R}[x_1, \ldots, x_m]$.

(i) If $\{g_1, \ldots, g_t\}$ is a (reduced Gröbner) basis for I (w.r.t. $\prec_m$), then $\{g_1, \ldots, g_t\}$ is a (reduced Gröbner) basis for I' (w.r.t. $\prec_n$).

(ii) Let $f = \prod_{k \in [m]} x_k \in \mathbb{R}[x_1, \ldots, x_m]$. If $I : f^\infty \subseteq \mathbb{R}[x_1, \ldots, x_m]$ has a basis $\{h_1, \ldots, h_l\}$, then $I' : f^\infty =< h_1, \ldots, h_l > \subseteq \mathbb{R}[x_1, \ldots, x_n]$.

Proof. (i) Since $\{g_1, \ldots, g_t\}$ is a basis for I, then $f_i = \sum_{j=1}^t h_j^{(i)} g_j, h_j \in \mathbb{R}[x_1, \ldots, x_m]$,
$i = 1, \ldots, s$. We note that $I' =< f_1, \ldots, f_s >$ and then $I' =< g_1, \ldots, g_t >$. Furthermore, if $\{g_1, \ldots, g_t\}$ is a reduced Gröbner basis for I, from Buchberger's algorithm and the definition of a reduced Gröbner basis, we have that $\{g_1, \ldots, g_t\}$ is a reduced Gröbner basis for I'.

(ii) Fix the grevlex monomial order $\prec_n$ induced by $x_n > \cdots > x_1$. $\prec_m$ is a monomial order derived from $\prec_n$. Since I and I' have the same reduced Gröbner basis, by Sturmfels (1996, Lemma 12.1). $I : x_1^\infty$ and $I' : x_1^\infty$ have the same reduced Gröbner basis. If we change the monomial order to a grevlex monomial order induced by $x_n > \cdots > x_3 > x_1 > x_2$, $(I : x_1^\infty) : x_2^\infty$ and $(I' : x_1^\infty) : x_2^\infty$ have the same reduced Gröbner basis as well. Considering $x_1, x_2, \ldots, x_m$ one by one, $I : f^\infty$ has the same basis as $I' : f^\infty$, and the conclusion follows.

Now we give the proof of Theorem 1.

Proof. Let $g_1, \ldots, g_t$ be polynomials arising from the pairwise Markov properties of $\mathbf{G}$. According to Geiger, Meek, and Sturmfels (2006),

$$< f_1, f_2, \ldots, f_l >=< g_1, \ldots, g_t >: (\prod_{\alpha \in \mathcal{S}} x_\alpha)^\infty.$$

Let $e \in [\gamma_{n+1}]$. Let $U_e = \{(d_1, \ldots, d_n, e) \,|\, d_i \in [\gamma_i], i \in [n]\}$ and $g_1^{(e)}, \ldots, g_t^{(e)} \in \mathbb{R}[\tilde{X}_{U_e}]$ be polynomials arising from the pairwise Markov properties of $\hat{\mathbf{G}}$. Suppose

$$I_e =< g_1^{(e)}, \ldots, g_t^{(e)} > \subseteq \mathbb{R}[\tilde{X}_{U_e}],\ \tilde{I}_e =< g_1^{(e)}, \ldots, g_t^{(e)} > \subseteq \mathbb{R}[\tilde{X}].$$

It is easy to verify that

$$\Psi_e : \mathbb{R}[X] \to \mathbb{R}[\tilde{X}_{U_e}],$$
$$\sum_\alpha a_\alpha X^\alpha \mapsto \sum_\alpha a_\alpha \tilde{X}_{U_e}^\alpha, \ \alpha \in \mathbb{Z}_+^{\gamma_1 \times \cdots \times \gamma_n}.$$

is an isomorphism.

By Sturmfels (1996, Lemma 12.1) and Corollary A.2,

$$I_{\hat{\mathbf{G}}} = (\sum_{e=1}^{\gamma_{n+1}} \tilde{I}_e) : (\prod_{\beta \ in \mathcal{T}} \tilde{x}_\beta)^\infty = \sum_{e=1}^{\gamma_{n+1}} (\tilde{I}_e : (\prod_{\beta \in U_e} \tilde{x}_\beta)^\infty).$$

Since $g_i^{(e)} \in \mathbb{R}[\tilde{X}_{U_e}], i \in [t]$, by Lemma A.1, a basis for $I_{U_e} = I_e : (\prod_{\beta \in U_e} \tilde{x}_\beta)^\infty \subseteq R[\tilde{X}_{U_e}]$ is a basis for $\tilde{I}_e : (\prod_{\beta \in U_e} \tilde{x}_\beta)^\infty \subseteq R[\tilde{X}]$. Then the basis for $\sum_{e=1}^{\gamma_{n+1}} I_{U_e}$ is a basis for $\sum_{e=1}^{\gamma_{n+1}} (\tilde{I}_e : (\prod_{\beta \in U_e} \tilde{x}_\beta)^\infty)$.

Now we compute a basis for I_{U_e}. Since Ψ_e is an isomorphism,

$$I_{U_e} = I_e : (\prod_{\beta \in U_e} \tilde{x}_\beta)^\infty = \Psi_e(I : X^\infty) =< f_1^{(e)}, f_2^{(e)}, \ldots, f_l^{(e)} > .$$

Then $\bigcup_{e=1}^{\gamma_{n+1}} < f_1^{(e)}, f_2^{(e)}, \ldots, f_l^{(e)} >$ is the generator set desired.

References

Cox, D., Little, J., and O'Shea, D. (2007). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.* 3rd edition, Springer, New York.

De Loera, J. and Onn, S. (2006). Markov bases for three-way tables are arbitrarily complicated. *J. Symb. Comput.* **41**, 173-181.

Dellaportas, P. and Tarantola, C. (2005). Model determination for categorical data with factor level merging. *J. Roy. Statist. Soc. Ser. B* **67**, 269-283.

Develin, M. and Sullivant, S. (2003). Markov bases for binary graph models. *Ann. Comb.* **7**, 441-466.

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26**, 363-397.

Dobra, A. (2003). Markov bases for decomposable graphical models. *Bernoulli* **9**, 1093-1108.

Dobra, A. and Sullivant, S. (2004). A divide-and-conquer algorithm for generating Markov bases of multi-way tables. *Comput. Statist.* **19**, 347-366.

Drton, M. and Sullivant, S. (2007). Algebraic statistical models. *Statist. Sinica* **17**, 1273-1297.

Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics.* Birkhäuser Verlag AG, Basel · Boston · Berlin.

Eisenbud, D. and Sturmfels, B. (1996). Binomial ideals. *Duke Math. J.* **84**, 1-45.

Fienberg, E. (2007). Expanding the statistical toolkit with algebraic statistics. *Statist. Sinica* **17**, 1261-1272.

Fienberg, E., Hersh, P., Rinaldo, A. and Zhou Y. (2010). Maximum likelihood estimation in latent class models for contingency tables. *Algebraic and Geometric Methods in Statistics* (Edited by Gibilisco, P., Riccomagno, E., Rogantin, M. and Wynn, H.). Cambridge University Press.

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

Haberman, S. J. (1988). A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *J. Amer. Statist. Assoc.* **83**, 555-560.

Hoşten, S. and Sullivant, S. (2002). Gröbner bases and polyhedral geometry of reducible and cyclic models. *J. Combin. Theory Ser. A.* **100**, 277–301.

Král', D., Norine, S. and Pangrác, O. (2010). Markov bases of binary graph models of K_4-minor free graphs. *J. Combin. Theory Ser. B.* **117**, 759-765.

Krampe, A. and Kuhnt, S. (2010). Model selection for contingency tables with algebraic statistics. *Algebraic and Geometric Methods in Statistics.* (Edited by Gibilisco, P., Riccomagno, E., Rogantin, M. and Wynn, H.). Cambridge University Press.

Lauritzen, S. L. (1996). *Graphical Models.* Clarendon Press.

Lauritzen, S. L. (2002). *Lectures on Contingency Tables.* Electronic edition.

Leimer, H.-G. (1993). Optimal decomposition by clique separators. *Discrete Math.* **113**, 99-123.

Riccomagno, E. (2009). A short history of algebraic statistics. *Metrika* **69**, 397-418.

Sturmfels, B. (1996). *Gröbner Bases and Convex Polytopes.* American Mathematical Society, Univ. Lectures Series, No 8, Providence.

Sturmfels, B. and Sullivant, S. (2008). Toric geometry of cuts and splits. *Michigan Math. J.* **57**, 689-709.

Tarjan, R. (1985), Decomposition by clique separators. *Discrete Math.* **55**, 221-232.

Wermuth, N. and Cox, D. (1998). On the application of conditional independence to ordinal data. *Int. Statist. Rev.* **66**, 181–199.

Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin Province, China.

Key Laboratory of Symbolic Computation and Knowledge Engineering of MOE, Jilin University, China.

E-mail: caisf@nenu.edu.cn

Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin Province, China.

E-mail: libc580@nenu.edu.cn

Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin Province, China.

E-mail: jhguo@nenu.edu.cn

(Received June 2011; accepted January 2013)