NISS

A Divide-and-Conquer Algorithm for Generating Markov Bases of Multi-way Tables

Adrian Dobra and Seth Sullivant

Technical Report Number 124
March, 2002

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

A Divide-and-Conquer Algorithm for Generating Markov Bases of Multi-way Tables

Adrian Dobra[1] and Seth Sullivant[2]

[1] National Institute of Statistical Sciences, Research Triangle Park, NC 27709
[2] Department of Mathematics, San Francisco State University, San Francisco, CA 94132

Summary

We describe a divide-and-conquer technique for generating a Markov basis that connects all tables of counts having a fixed set of marginal totals. This procedure is based on deposing the independence graph induced by these marginals. We discuss the practical imports of using this method in conjunction with other algorithms for determining Markov bases.

Keywords: Contingency tables; Decomposable graphical models; Disclosure limitation; Exact distributions; Gröbner bases; Markov bases

1 Introduction

The focus of this paper are sets of k-way tables of counts that are induced by fixing several lower dimensional marginals. These sets arise in a variety of contexts such as disclosure limitation (Dobra 2002, Fienberg, Makov & Steele 1998, Fienberg, Makov, Meyer & Steele 2001) and the calibration of test statistics (Agresti 1992, Diaconis & Efron 1985, Mehta 1994). Although there exist efficient algorithms for counting and enumerating two-way integer tables with fixed row and column totals (Mount 1995, De Loera & Sturmfels 2001), the problem of describing sets of k-way tables, $k > 2$, remains open in the literature. Diaconis and Sturmfels (1998) proposed a general algorithm for generating random draws from a set of tables with given fixed marginals. Their approach is extremely appealing because, in theory, it can be used for arrays of any dimension. Despite its generality, the power of this sampling procedure is limited because it requires access to a Markov basis–a finite set of data swaps which allow any two tables with the same fixed marginals to be connected. In addition to sampling, Markov bases can be employed to enumerate all the integer tables with a given set of marginals. As a consequence, Markov bases allow one to create a "replacement" for a database consisting of a k-way contingency table, when such a replacement is needed to protect the individuals with rare characteristics whose identity might be disclosed by the release of a number of marginals (Willenborg & de Waal 2000).

Diaconis and Sturmfels (1998) and Dinwoodie (1998) suggest computing a Markov basis by finding a Gröbner basis (Cox, Little & O'Shea 1992) of a well-specified polynomial ideal, but their method is difficult to employ even for tables with three dimensions because of the computational complexity of computing Gröbner bases. The statistical theory on graphical models (Madigan & York 1995, Whittaker 1990, Lauritzen 1996) shows that the conditional dependencies induced by a set of fixed marginals among the variables cross-classified in a table of counts can be visualized by means of an independence graph. In particular, a lot of attention has been given to *decomposable* graphs (Lauritzen 1996), a special class of graphs that can be "broken" into components such that (i) every component is associated with exactly one fixed marginal; and (ii) no information is lost in the decomposition process, i.e. no marginal is "split" between two components. *Reducible* graphs (Tarjan 1985, Leimer 1993) are generalizations of decomposable graphs. A reducible graph is one that can be at least partially decomposed, though the resulting components of the decomposition may correspond to more than one fixed marginal.

In a companion paper, Dobra (2001) gave explicit formulas for dynamically generating a Markov basis when the set of fixed marginals defines a decomposable independence graph. By following a similar line of reasoning, we develop a divide-and-conquer algorithm for significantly reducing the time needed to find a Markov basis when the underlying independence graph is not decomposable, but is reducible. In this case, the problem of finding a Markov basis can be reduced to finding Markov bases corresponding to smaller components of the graph. The major advantage of using our algorithm comes from the fact that the number of variables involved is one of the main factors that leads to an exponential increase in the amount of computations required for generating a Markov basis. We show that "breaking" a set of k-way tables into several sets of tables of dimension strictly smaller than k, and "assembling" the resulting Markov bases associated with these lower-dimensional families of tables can be done *at almost no computational cost*.

In the next section, we outline the basic theory of tables and Markov bases. In Section 3 we introduce decomposable and reducible sets of marginals and discuss some of their properties. In Section 4 we outline the divide-and-conquer technique for generating Markov bases. In Section 5 we prove the main theorem of the paper which constitutes the core of our dimension reduction algorithm. In the final section we summarize our results and draw some conclusions.

2 Data Swapping and Markov Bases

A k-way table of counts $\mathbf{n}_K$ is a k-dimensional array of non-negative integer numbers. Each variable X_j, $j = 1, 2, \ldots, k$, recorded in such a table can take a finite number of values $x_j \in \mathcal{I}_j := \{1, 2, \ldots, I_j\}$. Let $\mathcal{I} = \mathcal{I}_1 \times \mathcal{I}_2 \times \ldots \times \mathcal{I}_k$. A *cell entry* $n_K(i_K)$, $i_K \in \mathcal{I}_K$, in table $\mathbf{n}_K$ is a non-negative integer representing the number of units or individuals sharing the same attributes i_K.

Let $D = \{i_1, i_2, \ldots, i_l\}$ denote an arbitrary subset of K. The D-marginal $\mathbf{n}_D$ of $\mathbf{n}_K$ is the contingency table with *marginal cells* $i_D \in \mathcal{I}_D := \mathcal{I}_{i_1} \times \ldots \times \mathcal{I}_{i_l}$ and entries given by

$$n_D(i_D) = \sum_{j_{K\setminus D} \in \mathcal{I}_{K\setminus D}} n_K(i_D, j_{K\setminus D}).$$

Two k-way tables $\mathbf{n}_K^1 = \left\{n_K^1(i_K)\right\}_{i_K \in \mathcal{I}_K}$ and $\mathbf{n}_K^2 = \left\{n_K^2(i_K)\right\}_{i_K \in \mathcal{I}_K}$, are *equal* if $n_K^1(i_K) = n_K^2(i_K)$ for all $i_K \in \mathcal{I}_K$, and in this case we write $\mathbf{n}_K^1 = \mathbf{n}_K^2$. If all the counts in table $\mathbf{n}_K^1$ are zero, i.e. $n_K^1(i_K) = 0$, for all $i_K \in \mathcal{I}_K$, we write $\mathbf{n}_K^1 = \mathbf{0}$. The *sum* of two k-way tables $\mathbf{n}_K^1$ and $\mathbf{n}_K^2$ is another k-way table $\mathbf{n}_K^3 := \mathbf{n}_K^1 + \mathbf{n}_K^2$ with entries $n_K^3(i_K) = n_K^1(i_K) + n_K^2(i_K)$, for $i_K \in \mathcal{I}_K$. Similarly, the difference between $\mathbf{n}_K^1$ and $\mathbf{n}_K^2$ is an array $\mathbf{n}_K^4 := \mathbf{n}_K^1 - \mathbf{n}_K^2$ with entries $n_K^4(i_K) = n_K^1(i_K) - n_K^2(i_K)$. The set of k-way tables indexed by $\mathcal{I}_K$ is also closed under multiplication with scalars. If a is a real/integer number, the array $\mathbf{n}_K^5 := a \cdot \mathbf{n}_K^1$ has entries $n_K^5(i_K) = a \cdot n_K^1(i_K)$.

Data swapping (Dalenius & Reiss 1982) is a disclosure limitation technique that involves moving table entries from one cell to the other. Since some of the cell entries could be increased and other cell entries could be decreased, a *data swap* associated with a k-way table $\mathbf{n}_K$ is a k-way array $\mathbf{f}_K = \{f_K(i_K)\}_{i_K \in \mathcal{I}_K}$ containing integer entries, i.e.

$$f_K(i_K) \in \{\ldots, -2, -1, 0, 1, 2, \ldots\},$$

for all $i_K \in \mathcal{I}_K$. Intuitively, a data swap can be viewed as the difference between the post-swapped and the pre-swapped tables. The table created by repeatedly applying data swaps to the original table is sometimes required to be consistent with the marginals that were previously made public (Willenborg & de Waal 2000). Consequently, we are interested in data swaps that leave a number of marginals unchanged.

Definition 2.1. Let D_1, D_2, ..., D_r be subsets of K. A *move* $\mathbf{f}_K$ for $D_1, \ldots, D_r$ is a data swap that preserves the marginal tables specified by the index sets D_1, D_2, ..., D_r. In other words, $\mathbf{f}_{D_j} = \mathbf{0}$, for all $j = 1, 2, \ldots, r$.

Let $\mathbf{e}^{i_K} = \left\{e^{i_K}(j_K)\right\}_{j_K \in \mathcal{I}_K}$ be a k-way table that has all cell entries equal to zero except the i_K-th, where it contains a count of "1", i.e.

$$e^{i_K}(j_K) = \begin{cases} 1, & \text{if } j_K = i_K, \\ 0, & \text{otherwise.} \end{cases}$$

With this notation, a move $\mathbf{f}_K = \{f_K(i_K)\}_{i_K \in \mathcal{I}_K}$ is given by:

$$\mathbf{f}_K = \sum_{i_K \in \mathcal{I}_K} f_K(i_K) \cdot \mathbf{e}^{i_K}.$$

Let $\mathcal{I}_+^{\mathbf{f}_K} := \{i_K \in \mathcal{I}_K : f_K(i_K) > 0\}$ and $\mathcal{I}_-^{\mathbf{f}_K} := \{i_K \in \mathcal{I}_K : f_K(i_K) < 0\}$ be the set of indices corresponding with the positive and the negative cell entries in $\mathbf{f}_K$. Then we can write

$$\mathbf{f}_K = \sum_{i_K \in \mathcal{I}_+^{\mathbf{f}_K}} f_K(i_K) \cdot \mathbf{e}^{i_K} + \sum_{i_K \in \mathcal{I}_-^{\mathbf{f}_K}} f_K(i_K) \cdot \mathbf{e}^{i_K}.$$

We can define the *multiset* $P(\mathbf{f}_K)$ to consist of all the indices i_K in $\mathcal{I}_+^{\mathbf{f}_K}$, but where i_K occurs $f_K(i_K)$ times. Similarly, define the multiset $N(\mathbf{f}_K)$ to consist of all indices i_K in $\mathcal{I}_-^{\mathbf{f}_K}$ with i_K occurring $-f_K(i_K)$ times. Now we can write $\mathbf{f}_K$ as

$$\mathbf{f}_K = \sum_{i_K \in P(\mathbf{f}_K)} \mathbf{e}^{i_K} - \sum_{i_K \in N(\mathbf{f}_K)} \mathbf{e}^{i_K}.$$

By definition, a move $\mathbf{f}_K$ should preserve *at least* one marginal total, hence the sum of all entries of $\mathbf{f}_K$ has to be zero. Therefore we need to have:

$$\sum_{i_K \in \mathcal{I}_+^{\mathbf{f}_K}} f_K(i_K) = - \sum_{i_K \in \mathcal{I}_-^{\mathbf{f}_K}} f_K(i_K).$$

This implies that the multisets $P(\mathbf{f}_K)$ and $N(\mathbf{f}_K)$ have the same cardinality, i.e. $P(\mathbf{f}_K) = \{i_1, \ldots, i_m\}$ and $N(\mathbf{f}_K) = \{i_1{}', \ldots, i_m{}'\}$ with $i_l \in \mathcal{I}_+^{\mathbf{f}_K}$, $i_l{}' \in \mathcal{I}_-^{\mathbf{f}_K}$ for $l = 1, \ldots, m$. We represent moves more compactly as:

$$\begin{aligned}\mathbf{f}_K &= [P(\mathbf{f}_K)||N(\mathbf{f}_K)],\\ &= [\{i_1,\dots,i_m\}||\{i_1{}',\dots,i_m{}'\}].\end{aligned}$$

Example 2.2. Let $K = \{1,2,3,4\}$ and $\mathcal{I}_1 = \mathcal{I}_2 = \mathcal{I}_3 = \mathcal{I}_4 = \{1,2\}$. A move $\mathbf{f}_K = \{f_K(i_K)\}_{i_K \in \mathcal{I}_K}$ can be represented as a nested 2-dimensional array:

$$\mathbf{f}_K = \begin{pmatrix} (f(i,j,1,1)) & (f(i,j,1,2)) \\ (f(i,j,2,1)) & (f(i,j,2,2)) \end{pmatrix}.$$

For example, the move

$$\mathbf{f}_K = \begin{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \\ \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix},$$

preserves the $\{1,2\}$, $\{2,3\}$ and $\{3,4\}$- marginals of a four-way table because the corresponding two-way marginals of $\mathbf{f}_K$ are zero. In our notation, we write $\mathbf{f}_K$ as:

$$\mathbf{f}_K = [\{(2,1,2,1),(1,1,1,2)\}||\{(2,1,1,2),(1,1,2,1)\}].$$

■

This notation can also be used for arbitrary non-negative integral tables. One advantage of using our notation is that, given $D \subset K$ and a table $\mathbf{n}_K$, it is easy to compute the D-marginal of $\mathbf{n}_K$.

Lemma 2.3. *Let $\mathbf{n}_K$ be a k-way table with non-negative entries. Suppose that $D = \{1,2,\dots,a\}$ and $K \setminus D = \{a+1,\dots,k\}$. We write $\mathbf{n}_K$ as*

$$\begin{aligned}\mathbf{n}_K &= [P(\mathbf{n}_K)||\emptyset],\\ &= [\{(s_1,t_1),\dots,(s_m,t_m)\}||\emptyset].\end{aligned}$$

All the s_j index $\mathcal{I}_D$ and the t_j index $\mathcal{I}_{K\setminus D}$. Then $\mathbf{n}_D$ is given by

$$\begin{aligned}\mathbf{n}_D &= [P(\mathbf{n}_D)||\emptyset],\\ &= [\{s_1,\dots,s_m\}||\emptyset].\end{aligned}$$

Proof. The D-marginal of $\mathbf{e}^{(s,t)}$ is $\mathbf{e}^s$. The lemma then follows by linearity of computing D-marginals. ■

Denote by $\mathbf{T}(\mathbf{n}_{D_1}, \ldots, \mathbf{n}_{D_r})$ the set of all tables $\mathbf{n}_K$ that have their D_1, D_2, ..., D_r-marginals equal to the corresponding marginals of $\mathbf{n}_K$. A move $\mathbf{f}_K$ is *admissible* for $\mathbf{n}_K$ if $\mathbf{n}_K + \mathbf{f}_K$ belongs to $\mathbf{T}(\mathbf{n}_{D_1}, \ldots, \mathbf{n}_{D_r})$. In particular, we must have $(n_K + f_K)(i_K) \geq 0$ for all indices i_K, which is equivalent to $N(\mathbf{f}_K) \subset P(\mathbf{n}_K)$.

Definition 2.4. A *Markov basis* $\mathcal{M}$ is a finite collection of moves that preserve the $D_1, \ldots, D_r$-marginals and connects any two k-way tables that have the same $D_1, \ldots, D_r$-marginals. In other words, for any two tables $\mathbf{x}_K$, $\mathbf{n}_K$ such that $\mathbf{x}_K \in \mathbf{T}(\mathbf{n}_{D_1}, \ldots, \mathbf{n}_{D_r})$, there exists a sequence of moves $\mathbf{f}_K^1$, $\mathbf{f}_K^2$, ..., $\mathbf{f}_K^s$ in $\mathcal{M}$ such that

$$\mathbf{x}_K - \mathbf{n}_K = \sum_{j=1}^{s} \mathbf{f}_K^j, \tag{2.1}$$

and

$$\mathbf{n}_K + \sum_{j=1}^{s'} \mathbf{f}_K^j \in \mathbf{T}(\mathbf{n}_{D_1}, \ldots, \mathbf{n}_{D_r}), \tag{2.2}$$

for $1 \leq s' \leq s$. Eq. 2.1 and Eq. 2.2 say that the table $\mathbf{n}_K$ is *transformed* into $\mathbf{x}_K$ by employing moves in $\mathcal{M}$. Since a Markov basis $\mathcal{M}$ depends only on the index sets $\mathcal{I}_{D_1}$, ..., $\mathcal{I}_{D_r}$, we will say that $\mathcal{M}$ is a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$, where

$$\mathbf{T}(D_1, \ldots, D_r) = \{\mathbf{T}(\mathbf{n}_{D_1}, \ldots, \mathbf{n}_{D_r}) : \mathbf{n}_K \text{ is a table of counts}\}.$$

Theorem 2.5. *(Diaconis & Sturmfels 1998) There exists a Markov basis $\mathcal{M}$ for $\mathbf{T}(D_1, \ldots, D_r)$.*

In the algebraic setting, a Markov basis for a set of tables corresponds to a *generating set* for a corresponding *toric ideal*. For detailed explanations, see Diaconis and Sturmfels (1998).

3 Special Configurations of Marginals

In this section we closely follow the notation and definitions relating to graph theory introduced in Lauritzen (1996) and Dobra and Fienberg (2000). A brief introduction with the basic graph terminology needed to understand the results that follow is given in Appendix A.

Consider a set of marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, ..., $\mathbf{n}_{D_r}$ such that their index sets cover K, i.e. $K = D_1 \cup D_2 \cup \ldots \cup D_r$. We always assume that there are no redundant configurations in this sequence, that is there are no r_1, r_2 such that $D_{r_1} \subset D_{r_2}$.

We visualize the dependency patterns induced by D_1, D_2, ..., D_r by constructing an independence graph. Each vertex in this graph represents a variable X_j, $j \in K$. We draw an edge between two vertices if and only if the two-dimensional array defined by the variables associated with these vertices is a marginal of some $\mathbf{n}_{D_l}$.

Definition 3.1. The *independence graph* $\mathcal{G} = \mathcal{G}(D_1, D_2, \ldots, D_r)$ associated with $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, ..., $\mathbf{n}_{D_r}$ is a graph with vertex set $K = D_1 \cup\ D_2 \cup \ldots \cup D_r$ and edge set E given by

$$E := \{(u, v) : \{u, v\} \subset D_j, \text{ for some } j \in \{1, \ldots, r\}\}.$$

Log-linear models are the usual way of representing and studying contingency tables with fixed marginals (Bishop 1975). If the minimal sufficient statistics of a log-linear model define a decomposable independence graph, the model is said to be *decomposable*. By analogy with log-linear models theory, we introduce decomposable sets of marginals.

Definition 3.2. The set of marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, ..., $\mathbf{n}_{D_r}$ is called *decomposable* if its corresponding independence graph $\mathcal{G} = \mathcal{G}(D_1, D_2, \ldots, D_r)$ is decomposable and the cliques $\mathcal{C}(\mathcal{G})$ of $\mathcal{G}$ are the index sets associated with $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, ..., $\mathbf{n}_{D_r}$, i.e.

$$\mathcal{C}(\mathcal{G}) = \{D_1, D_2, \ldots, D_r\}.$$

Therefore a decomposable set of marginals could represent the minimal sufficient statistics of a decomposable log-linear model.

Example 3.3. Let $\mathbf{n}_K$ be a table of counts cross-classifying variables X_1, X_2,..., X_6. With every variable X_j we associate a vertex $j \in K = \{1, \ldots, 6\}$. The index sets $D_1 = \{2, 6\}$, $D_2 = \{1, 2, 3, 5\}$, $D_3 = \{1, 4, 5\}$ define a two-way, a four-way and a three-way marginal of $\mathbf{n}_K$, respectively. The independence graph $\mathcal{G}$ induced by $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$ and $\mathbf{n}_{D_3}$ is represented in Fig. 1. If we "break" this graph in the vertex "2" and along the edge "$(1, 5)$", we end up with three components corresponding to the marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$ and $\mathbf{n}_{D_3}$. Any of these components cannot be further "broken". As a consequence, $\mathcal{G}$ is a decomposable graph with cliques $\mathcal{C}(\mathcal{G}) = \{D_1, D_2, D_3\}$, while $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$ and $\mathbf{n}_{D_3}$ form a decomposable set of marginals. The separators of $\mathcal{G}$ are $S_2 = \{2\}$ and $S_3 = \{1, 5\}$. ■

Definition 3.4. The marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, ..., $\mathbf{n}_{D_r}$ are *consistent* if, for any r_1, r_2, the $(D_{r_1} \cap D_{r_2})$-marginal of $\mathbf{n}_{D_{r_1}}$ is equal to the $(D_{r_1} \cap D_{r_2})$-marginal of $\mathbf{n}_{D_{r_2}}$.

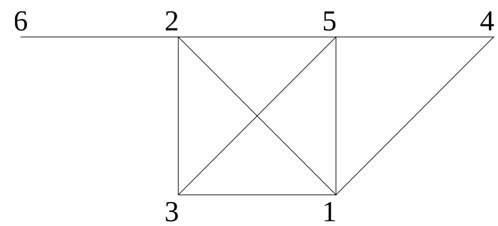

Figure 1: An example of a decomposable independence graph induced by a two-way, a four-way and a three-way marginal of a six-way contingency table.

The consistency of a set of marginals does not necessarily imply the existence of a table having this particular set of marginal totals–see, for example, Vlach (1986). To be more precise, $\mathbf{T}(\mathbf{n}_{D_1},\ldots,\mathbf{n}_{D_r})$ could be empty even if $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, …, $\mathbf{n}_{D_r}$ are consistent. In the special case of consistent and decomposable marginals, however, $\mathbf{T}(\mathbf{n}_{D_1},\ldots,\mathbf{n}_{D_r})$ is never empty (Dobra 2002).

Decomposable log-linear models have many other exceptional properties that have been well documented in the literature. For example, the corresponding maximum likelihood estimates can be expressed in closed form (Lauritzen 1996, Whittaker 1990). Additionally, Dobra and Fienberg (2000) obtain formulas for calculating sharp upper and lower bounds for cell entries of tables in $\mathbf{T}(\mathbf{n}_{D_1},\ldots,\mathbf{n}_{D_r})$ given that the marginals are consistent and decomposable.

Decomposable sets of marginals also have extremely simple Markov bases. Dobra (2001) constructs a Markov basis $\mathcal{F}(D_1,\ldots,D_r)$ for $\mathbf{T}(D_1,\ldots,D_r)$ if $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, …, $\mathbf{n}_{D_r}$ is a decomposable set of marginals. All the moves in these Markov bases are *primitive*, i.e. they have two entries equal to "1", two entries equal to "-1", while the remaining cells are zero. For related efforts for determining a Markov basis in the decomposable case see Takken (1999).

If a configuration of marginals is not decomposable, it seems natural to ask whether we could reduce the computational effort needed to generate a Markov basis by employing the same strategy used in the decomposable case: decompositions of graphs by means of separators. We explore the more general situation when the independence graph $\mathcal{G}$ is not decomposable, but is reducible.

Definition 3.5. The set of tables is $\mathbf{n}_{D_1}, \mathbf{n}_{D_2}, \ldots, \mathbf{n}_{D_r}$ called *reducible* if its corresponding independence graph $\mathcal{G} = \mathcal{G}(D_1, D_2, \ldots, D_r)$ has a decomposition (A_1, S, A_2) with $S \subset D_l$ for some l.

For example, any decomposable set of marginals is also reducible but the converse is not necessarily true. Additionally, it may happen that the underlying independence graph is reducible although the set of marginals is not.

Example 3.6. Let $K = \{1,2,3,4,5\}$ and $\mathbf{n}_K$ be a five-way table. Consider the index sets $D_1 = \{1,2,3\}$, $D_2 = \{1,3,4\}$, $D_3 = \{2,3,4\}$, $D_4 = \{1,4,5\}$, $D_5 = \{2,4,5\}$, and $D_6 = \{1,2,5\}$. Notice that $D_1 \cup D_2 \cup \ldots \cup D_6 = K$. The independence graph $\mathcal{G}$ defined by these index sets is pictured in Fig. 2. The cliques of $\mathcal{G}$ are $V_1 := \{1,2,3,4\}$ and $V_2 := \{1,2,4,5\}$. Because D_1, D_2, …, D_6 are not cliques in $\mathcal{G}$, the set of marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, …, $\mathbf{n}_{D_6}$ is not decomposable.

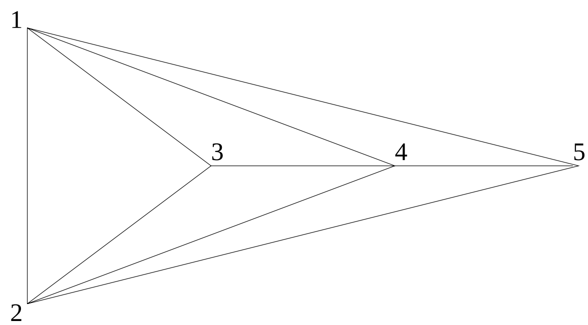

Figure 2: The independence graph induced by six three-way marginals of a five-way table. Although the graph is decomposable, this set of marginals is neither decomposable nor reducible.

Take $S := V_1 \cap V_2 = \{1,2,4\}$. Since $(V_1 \setminus S, S, V_2 \setminus S)$ is a proper decomposition of $\mathcal{G}$, this graph is decomposable hence reducible. However, the set $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, …, $\mathbf{n}_{D_6}$ is *not* reducible because there is no D_l, $l \in \{1,2,\ldots,6\}$, such that $S \subset D_l$. ■

4 Divide-and-Conquer

The special decomposability properties of independence graphs can be exploited to significantly reduce the computational effort required to generate a Markov basis $\mathbf{T}(D_1,\ldots,D_r)$ associated with a reducible set of marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, …, $\mathbf{n}_{D_r}$. Leimer (1993) shows that there exists a sequence of vertex sets V_1, V_2, …, V_q of the independence graph $\mathcal{G} = \mathcal{G}(D_1, D_2, \ldots, D_r) = (K, E)$ such that

$$K = V_1 \cup V_2 \cup \ldots \cup V_q,$$

and, for $j = 2,\ldots,q$, $(H_{j-1} \setminus S_j, S_j, V_j \setminus S_j)$, is a decomposition of the subgraph $\mathcal{G}(H_j)$ of $\mathcal{G}$. We denoted $H_j = V_1 \cup \ldots \cup V_j$ and $S_j = H_{j-1} \cap V_j$.

We call $S_2, \ldots, S_q$ the sequence of separators associated with $V_1, V_2, \ldots, V_q$. With every vertex set V_j, we associate

$$\mathcal{L}(V_j) := \{D_l : D_l \subset V_j\}. \tag{4.1}$$

Let $\mathcal{H}_j$ be a Markov basis for the class of tables $\mathbf{T}(\{D_l : D_l \in \mathcal{L}(V_j)\})$, for $j = 1, 2, \ldots, q$. Starting from $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_q$, we give a procedure for recursively generating a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$. Using direct algebraic calculations, a set of Markov bases $\{\mathcal{H}_j\}_j$ is easier to generate than a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$ since generating every $\mathcal{H}_j$ involves computations on a smaller set of variables $(X_i : i \in V_j)$.

We make the assumption that, for every separator S_j of $\mathcal{G}$, there exists some D_l, $l = 1, 2, \ldots, r$, such that $S_j \subset D_l$. It follows that, for $j = 2, \ldots, q$, the set of tables

$$\{\mathbf{n}_{D_l} : D_l \in \mathcal{L}(V_1) \cup \ldots \cup \mathcal{L}(V_j)\}, \tag{4.2}$$

is reducible. In the next section we describe a procedure for combining a Markov basis for the class of tables $\mathbf{T}(\mathcal{L}(V_1) \cup \ldots \cup \mathcal{L}(V_{j-1}))$ with a Markov basis $\mathcal{H}_j$ for $\mathbf{T}(\mathcal{L}(V_j))$ into a Markov basis for $\mathbf{T}(\mathcal{L}(V_1) \cup \ldots \cup \mathcal{L}(V_{j-1}) \cup \mathcal{L}(V_j))$. Given that such a procedure exists, we outline below an algorithm for determining a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$.

Algorithm 4.1 (Markov basis for a reducible set of marginals).

- *Obtain a Markov basis $\mathcal{H}^{1,2}$ for $\mathbf{T}(\mathcal{L}(V_1) \cup \mathcal{L}(V_2))$ from the Markov bases $\mathcal{H}_1$ and $\mathcal{H}_2$.*
- **for** $j = 3, \ldots, q$ **do**
  - *Obtain a Markov basis $\mathcal{H}^{1,2,\ldots,j}$ for $\mathbf{T}(\mathcal{L}(V_1) \cup \ldots \cup \mathcal{L}(V_j))$ from the Markov bases $\mathcal{H}^{1,2,\ldots,j-1}$ and $\mathcal{H}_j$.*

  end for

■

The resulting set of moves $\mathcal{H}^{1,2,\ldots,q}$ will be a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$.

Example 4.2. Let $K = \{1, 2, \ldots, 11\}$ and $\mathbf{n}_K$ be an eleven-way table. The index sets corresponding to the fixed marginals of $\mathbf{n}_K$ are the set of edges of the graph $\mathcal{G}$ in Fig. 3 from which we take out $\{3,4\}$, $\{3,11\}$, $\{4,11\}$, and then add $\{3,4,11\}$. $\mathcal{G}$ is a reducible graph and take $V_1 := \{2,3,9,10\}$, $V_2 := \{4,5,6,7\}$, $V_3 := \{1,3,4,11\}$ and $V_4 := \{3,4,7,8,9,11\}$. The separators

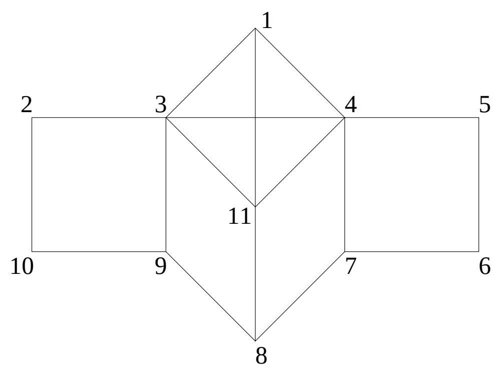

Figure 3: A reducible independence graph with four components (atoms) defined by the separators $\{3, 9\}$, $\{4, 7\}$ and $\{3, 4, 11\}$.

corresponding with $V_1, \ldots, V_4$ are $S_2 := \{3, 9\}$, $S_3 := \{4, 7\}$, $S_4 := \{3, 4, 11\}$. Because every separator identifies a fixed marginal, the set of tables $\mathbf{n}_{\{1,3\}}$, $\mathbf{n}_{\{1,11\}}$, $\mathbf{n}_{\{1,4\}}$, $\mathbf{n}_{\{3,4,11\}}$, $\mathbf{n}_{\{3,9\}}$, $\mathbf{n}_{\{8,11\}}$, $\mathbf{n}_{\{4,7\}}$, $\mathbf{n}_{\{8,9\}}$, $\mathbf{n}_{\{7,8\}}$, $\mathbf{n}_{\{2,3\}}$, $\mathbf{n}_{\{2,10\}}$, $\mathbf{n}_{\{9,10\}}$, $\mathbf{n}_{\{4,5\}}$, $\mathbf{n}_{\{5,6\}}$, $\mathbf{n}_{\{6,7\}}$ is reducible.

Assume we somehow managed to compute the Markov bases $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$, $\mathcal{H}_4$ associated with the classes of tables

$$\begin{aligned}
&\mathbf{T}(\{1,3\},\{1,11\},\{1,4\},\{3,4,11\}),\\
&\mathbf{T}(\{3,4,11\},\{3,9\},\{8,11\},\{4,7\},\{8,9\},\{7,8\}),\\
&\mathbf{T}(\{2,3\},\{3,9\},\{9,10\},\{2,10\}),\\
&\mathbf{T}(\{4,5\},\{5,6\},\{6,7\},\{4,7\}).
\end{aligned}$$

Then Algorithm 4.1 goes as follows:

- Make use of $\mathcal{H}_1$ and $\mathcal{H}_2$ to obtain a Markov basis $\mathcal{H}^{1,2}$ for

$$\mathbf{T}(\{1,3\},\{1,11\},\{1,4\},\{3,4,11\},\{3,9\},\{8,11\},\{4,7\},\{8,9\},\{7,8\}).$$

- Use $\mathcal{H}^{1,2}$ and $\mathcal{H}_3$ to determine a Markov basis $\mathcal{H}^{1,2,3}$ for

$$\mathbf{T}(\{1,3\},\{1,11\},\{1,4\},\{3,4,11\},\{3,9\},\{8,11\},\{4,7\},\{8,9\},\{7,8\},\\ \{2,3\},\{2,10\},\{9,10\}).$$

- Finally, combine $\mathcal{H}^{1,2,3}$ and $\mathcal{H}_4$ into a Markov basis $\mathcal{H}^{1,2,3,4}$ for

$$\mathbf{T}(\{1,3\},\{1,11\},\{1,4\},\{3,4,11\},\{3,9\},\{8,11\},\{4,7\},\{8,9\},\{7,8\},\\ \{2,3\},\{2,10\},\{9,10\},\{4,5\},\{5,6\},\{6,7\}).$$

■

5 A Decomposition Theorem

Consider a reducible set of marginals $\mathbf{n}_{D_1}$, $\mathbf{n}_{D_2}$, ..., $\mathbf{n}_{D_r}$ with independence graph $\mathcal{G} = (K, E)$. The graph $\mathcal{G}$ has a decomposition (A_1, S, A_2) such that $S \subset D_{l_0}$ for some $l_0 \in \{1, \ldots, r\}$. Without loss of generality, we can assume that $A_1 = \{1, \ldots, a\}$, $S = \{a+1, \ldots, b\}$, and $A_2 = \{b+1, \ldots, k\}$. Let $V_1 := A_1 \cup S$ and $V_2 := S \cup A_2$. Denote by $\mathcal{L}(V_1)$ and $\mathcal{L}(V_2)$ the set of all D_l, $l = 1, \ldots, r$, belonging to V_1 and V_2, respectively–see Eq. 4.1. Since S is a minimal separator of A_1 and A_2 in $\mathcal{G}$, the graph $\mathcal{G}'$ having cliques V_1 and V_2 is decomposable. Dobra (2001) constructs a set of moves $\mathcal{F}(V_1, V_2)$ for the class of tables $\mathbf{T}(V_1, V_2)$. This set is defined as follows:

Definition 5.1. $\mathcal{F}(V_1, V_2)$ is the union of all moves $\mathbf{f}_K$ and $-\mathbf{f}_K$, where $\mathbf{f}_K$ is given by

$$\begin{aligned} \mathbf{f}_K &= [P(\mathbf{f}_K)||N(\mathbf{f}_K)], \\ &= [\{(s_1,t,u_1),(s_2,t,u_2)\}||\{(s_1,t,u_2),(s_2,t,u_1)\}]. \end{aligned} \tag{5.1}$$

In Eq. 5.1 $s_1 \neq s_2$ range over $\mathcal{I}_{A_1}$, t ranges over $\mathcal{I}_S$, and $u_1 \neq u_2$ range over $\mathcal{I}_{A_2}$.

If $\mathbf{f}_K$ is as in Eq. 5.1, then $P(\mathbf{f}_{V_1}) = \{(s_1, t), (s_2, t)\} = N(\mathbf{f}_{V_1})$, which implies that $\mathbf{f}_{V_1} = \mathbf{0}$. Similarly, $\mathbf{f}_{V_2} = \mathbf{0}$, and consequently $\mathcal{F}(V_1, V_2)$ is indeed a set of moves for $\mathbf{T}(V_1, V_2)$.

Lemma 5.2. *(Dobra 2001) $\mathcal{F}(V_1, V_2)$ is a Markov basis for the class of tables $\mathbf{T}(V_1, V_2)$.*

Suppose we constructed a Markov basis $\mathcal{H}_1$ for $\mathbf{T}(\mathcal{L}(V_1))$ and a Markov basis $\mathcal{H}_2$ for $\mathbf{T}(\mathcal{L}(V_2))$. The goal of this section is to show how to generate a Markov basis $\mathcal{H}$ for $\mathbf{T}(D_1, \ldots, D_r)$ from $\mathcal{H}_1$ and $\mathcal{H}_2$. The first step is to "extend" the moves in $\mathcal{H}_1$ to a set of moves $\mathcal{H}_1'$ for $\mathbf{T}(D_1, \ldots, D_r)$. Similarly, the moves in $\mathcal{H}_2$ are extended to a set of moves $\mathcal{H}_2'$ for $\mathbf{T}(D_1, \ldots, D_r)$. In Theorem 5.6 we prove that

$$\mathcal{H}_1' \cup \mathcal{H}_2' \cup \mathcal{F}(V_1, V_2), \tag{5.2}$$

is a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$.

Let $\mathbf{f}_{V_1} = [P(\mathbf{f}_{V_1})||N(\mathbf{f}_{V_1})]$ be a move for $\mathbf{T}(\mathcal{L}(V_1))$, where

$$\begin{aligned} P(\mathbf{f}_{V_1}) &= \{(s_1, t_1), \ldots, (s_m, t_m)\}, \\ N(\mathbf{f}_{V_1}) &= \{(s_1{}', t_1{}'), \ldots, (s_m{}', t_m{}')\}, \end{aligned}$$

with the s_j and $s_j{}'$ indexing A_1 and the t_j and $t_j{}'$ indexing S. Since $S \subset D_{l_0}$, we must have $\mathbf{f}_S = \mathbf{0}$, but this implies that $P(\mathbf{f}_D) = N(\mathbf{f}_D)$ as multisets. Then there is a permutation and relabeling of the $\{(s_j{}', t_j{}')\}_{j=1}^m$ so that

$$\begin{aligned} \mathbf{f}_{V_1} &= [P(\mathbf{f}_{V_1})||N(\mathbf{f}_{V_1})], \\ &= [\{(s_1, t_1), \ldots, (s_m, t_m)\}||\{(s_1{}', t_1), \ldots, (s_m{}', t_m)\}]. \end{aligned} \tag{5.3}$$

Similarly, a move $\mathbf{f}'_{V_2}$ in $\mathbf{T}(\mathcal{L}(V_2))$ can be written as

$$\begin{aligned} \mathbf{f}'_{V_2} &= [P(\mathbf{f}'_{V_2})||N(\mathbf{f}'_{V_2})], \\ &= [\{(t_1, u_1), \ldots, (t_m, u_m)\}||\{(t_1, u_1{}'), \ldots, (t_m, u_m{}')\}]. \end{aligned}$$

Definition 5.3. Let $\mathbf{f}_{V_1} \in \mathcal{H}_1$ as in Eq. 5.3. Let $\mathbf{u} = \{u_1, \ldots, u_m\}$ be any set of m index sets from $\mathcal{I}_{A_2}$. Consider a map ψ that assigns to $\mathbf{f}_{V_1}$ and $\mathbf{u}$ a move $\mathbf{g}_K = \psi(\mathbf{f}_{V_1}, \mathbf{u})$ defined by

$$\begin{aligned} \mathbf{g}_K &= [P(\mathbf{g}_K)||N(\mathbf{g}_K)], \\ &= [\{(s_1, t_1, u_1), \ldots, (s_m, t_m, u_m)\}||\{(s_1{}', t_1, u_1), \ldots, (s_m{}', t_m, u_m)\}]. \end{aligned}$$

We define $\mathrm{Ext}(\mathcal{H}_1 \to \mathbf{T}(D_1, \ldots, D_r))$ as the union of all $\psi(\mathbf{f}_{V_1}, \mathbf{u})$ as $\mathbf{f}_{V_1}$ ranges over all moves in $\mathcal{H}_1$ and $\mathbf{u}$ ranges over all possible sequences of indices in $\mathcal{I}_{A_2}$.

Lemma 5.4. *If $\mathcal{H}_1$ is a set of moves for $\mathbf{T}(\mathcal{L}(V_1))$, $\mathrm{Ext}(\mathcal{H}_1 \to \mathbf{T}(D_1, \ldots, D_r))$ is a set of moves for $\mathbf{T}(D_1, \ldots, D_r)$.*

Proof. Let $\mathbf{g}_K = \psi(\mathbf{f}_{V_1}, \mathbf{u}) \in \mathrm{Ext}(\mathcal{H}_1 \to \mathbf{T}(D_1, \ldots, D_r))$. We must show that $\mathbf{g}_{D_l} = \mathbf{0}$, for all $l \in \{1, \ldots, r\}$. First, note that $N(\mathbf{g}_{V_2}) = P(\mathbf{g}_{V_2})$ hence $\mathbf{g}_{D_l} = \mathbf{0}$ for any $D_l \subset V_2$. On the other hand, $\mathbf{g}_{V_1} = \mathbf{f}_{V_1} \in \mathcal{H}_1$ is a move for $\mathbf{T}(\mathcal{L}(V_1))$. This implies that $\mathbf{g}_{D_l} = \mathbf{0}$ whenever $D_l \subset V_1$. Since any D_l is either a subset of V_1 or V_2, this proves the result. ■

Any move in $\mathcal{H}_2$ can be extended in an analogous way to moves for the class of tables $\mathbf{T}(D_1, \ldots, D_r)$, and the resulting set of moves will be denoted by $\mathrm{Ext}(\mathcal{H}_2 \to \mathbf{T}(D_1, \ldots, D_r))$.

Example 5.5. Let $K = \{1, 2, 3\}$ and $\mathbf{n}_K$ be a $2 \times 2 \times 2$ table with fixed one-way marginals $\mathbf{n}_{\{1\}}$, $\mathbf{n}_{\{2\}}$, and $\mathbf{n}_{\{3\}}$. The corresponding independence graph $\mathcal{G} = \mathcal{G}(\{1\}, \{2\}, \{3\})$ has three vertices and no edges. This graph is reducible with decomposition $(A_1, S, A_2) = (\{1, 2\}, \emptyset, \{3\})$. We have $V_1 = A_1 \cup S = \{1, 2\}$ and $V_2 = S \cup A_2 = \{3\}$. In addition, $\mathcal{L}(V_1) = \{\{1\}, \{2\}\}$ and

$\mathcal{L}(V_2) = \{\{3\}\}$. A Markov basis $\mathcal{H}_1$ for $\mathbf{T}(\{1\},\{2\})$ consists of the move $\mathbf{f}_{V_1}$ and its negative $-\mathbf{f}_{V_1}$, where $\mathbf{f}_{V_1}$ is given by

$$\begin{aligned} \mathbf{f}_{V_1} &= [P(\mathbf{f}_{V_1})||N(\mathbf{f}_{V_1})], \\ &= [\{(1,1),(2,2)\}||\{(1,2),(2,1)\}]. \end{aligned}$$

The set $\mathrm{Ext}(\mathcal{H}_1 \to \mathbf{T}(\{1\},\{2\},\{3\}))$ consists of the four moves of the form $\pm\mathbf{g}_K = \pm\psi(\mathbf{f}_{V_1},\{u_1,u_2\})$, with $u_1,u_2 \in \{1,2\}$ and

$$\begin{aligned} \mathbf{g}_K &= [P(\mathbf{g}_K)||N(\mathbf{g}_K)], \\ &= [\{(1,1,u_1),(2,2,u_2)\}||\{(1,2,u_1),(2,1,u_2)\}]. \end{aligned}$$

■

We are now ready to prove the main theorem of the paper.

Theorem 5.6. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Markov bases for $\mathbf{T}(\mathcal{L}(V_1))$ and $\mathbf{T}(\mathcal{L}(V_2))$, respectively. Then the set of moves*

$$\mathcal{H} = \mathrm{Ext}(\mathcal{H}_1 \to \mathbf{T}(D_1,\ldots,D_r)) \cup \mathrm{Ext}(\mathcal{H}_2 \to \mathbf{T}(D_1,\ldots,D_r)) \cup \mathcal{F}(V_1,V_2)$$

is a Markov basis for $\mathbf{T}(D_1,\ldots,D_r)$.

Proof. Let $\mathbf{x}_K^1$ and $\mathbf{x}_K^2$ be two k-way tables that belong to

$$\mathbf{T}(\mathbf{n}_{D_1},\ldots,\mathbf{n}_{D_r}) \in \mathbf{T}(D_1,\ldots,D_r).$$

We want to show that there exists a sequence of moves in $\mathcal{H}$ that connects $\mathbf{x}_K^1$ and $\mathbf{x}_K^2$. We can write

$$\begin{aligned} \mathbf{x}_K^1 &= [P(\mathbf{x}_K^1)||\emptyset], \\ &= [\{(s_1,t_1,u_1),\ldots,(s_m,t_m,u_m)\}||\emptyset], \end{aligned}$$

and

$$\begin{aligned} \mathbf{x}_K^2 &= [P(\mathbf{x}_K^2)||\emptyset], \\ &= [\{(s_1{}',t_1,u_1{}'),\ldots,(s_m{}',t_m,u_m{}')\}||\emptyset]. \end{aligned}$$

Both $\mathbf{x}_{V_1}^1$ and $\mathbf{x}_{V_1}^2$ belong to $\mathbf{T}(\{\mathbf{n}_{D_l} : D_l \in \mathcal{L}(V_1)\})$, hence there exists a sequence of moves $\mathbf{f}_{V_1}^1$, …, $\mathbf{f}_{V_1}^q$ in $\mathcal{H}_1$ that transforms $\mathbf{x}_{V_1}^1$ into $\mathbf{x}_{V_1}^2$. After a permutation and relabeling of the elements in $P(\mathbf{x}_K^1)$ we may suppose that $\mathbf{f}_{V_1}^1$ can be written as

$$\begin{aligned} \mathbf{f}_{V_1}^1 &= [P(\mathbf{f}_{V_1}^1)||N(\mathbf{f}_{V_1}^1)], \\ &= [\{(s_1{}'',t_1),\ldots,(s_j{}'',t_j)\}||\{(s_1,t_1),\ldots,(s_j,t_j)\}]. \end{aligned}$$

We "lift" this move to $\psi(\mathbf{f}^1_{V_1}, \mathbf{u}) \in \mathrm{Ext}(\mathcal{H}_1 \rightarrow \mathbf{T}(D_1, \ldots, D_r))$ with $\mathbf{u} = \{u_1, \ldots, u_j\}$, where the $\{u_1, \ldots, u_j\}$ come from $P(\mathbf{x}^1_K)$. Note that $\psi(\mathbf{f}^1_{V_1}, \mathbf{u})$ is an admissible move for $\mathbf{x}^1_K$ since $N(\psi(\mathbf{f}^1_{V_1}, \mathbf{u})) \subset P(\mathbf{x}^1_K)$. Repeating this process with all $\mathbf{f}_{V_1} \in \{\mathbf{f}^1_{V_1}, \ldots, \mathbf{f}^q_{V_1}\}$, we arrive at a sequence of moves in $\mathrm{Ext}(\mathcal{H}_1 \rightarrow \mathbf{T}(D_1, \ldots, D_r))$ that connects $\mathbf{x}^1_K$ to a table $\mathbf{x}^3_K$ with the property that $\mathbf{x}^3_{V_1} = \mathbf{x}^2_{V_1}$ and $\mathbf{x}^3_{V_2} = \mathbf{x}^1_{V_2}$.

Now, both $\mathbf{x}^3_{V_2}$ and $\mathbf{x}^2_{V_2}$ belong to $\mathbf{T}(\{\mathbf{n}_{D_l} : D_l \in \mathcal{L}(V_2)\})$. Then there exists a sequence of moves in $\mathcal{H}_2$ which connects $\mathbf{x}^3_{V_2}$ to $\mathbf{x}^2_{V_2}$. By a procedure analogous to the one in the preceding paragraph, this sequence of moves can be "lifted" to a sequence of moves in $\mathrm{Ext}(\mathcal{H}_2 \rightarrow \mathbf{T}(D_1, \ldots, D_r))$ that will transform $\mathbf{x}^3_K$ into a table $\mathbf{x}^4_K$ with the property that $\mathbf{x}^4_K$ belongs to the class of tables $\mathbf{T}(\mathbf{x}^2_{V_1}, \mathbf{x}^2_{V_2})$. Lemma 5.2 says that there is a sequence of moves in $\mathcal{F}(V_1, V_2)$ connecting $\mathbf{x}^4_K$ and $\mathbf{x}^2_K$. Putting together the three sequences of moves thus far described gives a sequence of moves connecting $\mathbf{x}^1_K$ and $\mathbf{x}^2_K$. ■

The next result is a straightforward consequence of Theorem 5.6.

Corollary 5.7. *Use the same notations as in Theorem 5.6. If $V_2 = D_l$ is a clique in the independence graph $\mathcal{G} = \mathcal{G}(D_1, \ldots, D_r)$ then*

$$\mathrm{Ext}(\mathcal{H}_1 \rightarrow \mathbf{T}(D_1, \ldots, D_r)) \cup \mathcal{F}(V_1, V_2),$$

is a Markov basis for $\mathbf{T}(D_1, \ldots, D_r)$.

Proof. In this case, $\mathcal{H}_2$ is the empty set and so $\mathrm{Ext}(\mathcal{H}_2 \rightarrow \mathbf{T}(D_1, \ldots, D_r))$ is empty also. The result then follows from Theorem 5.6. ■

Example 5.8. Let $K = \{1, 2, 3, 4\}$ and $\mathbf{n}_K$ be a $2 \times 2 \times 2 \times 2$ table with fixed marginals $\mathbf{n}_{\{1,2\}}$, $\mathbf{n}_{\{1,3\}}$, $\mathbf{n}_{\{2,3\}}$, $\mathbf{n}_{\{2,4\}}$, and $\mathbf{n}_{\{3,4\}}$. This set of marginals is reducible because the corresponding independence graph has a decomposition $(\{1\}, \{2, 3\}, \{4\})$. We can replace the problem of finding a Markov basis for this set of four-way tables with the problem of finding the Markov basis for two three-way tables. Take $V_1 = \{1, 2, 3\}$ and $V_2 = \{2, 3, 4\}$. A brute-force algebraic computation shows that a Markov basis $\mathcal{H}_1$ for $\mathbf{T}(\{1, 2\}, \{1, 3\}, \{2, 3\})$ consists of the moves $\pm \mathbf{f}_{V_1}$ with $\mathbf{f}_{V_1} = [P(\mathbf{f}_{V_1}) || N(\mathbf{f}_{V_1})]$, where

$$\begin{aligned} P(\mathbf{f}_{V_1}) &= \{(1,1,1), (1,2,2), (2,1,2), (2,2,1)\}, \text{ and} \\ N(\mathbf{f}_{V_1}) &= \{(2,1,1), (2,2,2), (1,1,2), (1,2,1)\}. \end{aligned}$$

The set of moves $\mathrm{Ext}(\mathcal{H}_1 \rightarrow \mathbf{T}(D_1, \ldots, D_r))$ consists of the sixteen moves of the form

$$\begin{aligned} \mathbf{g}_K &= \psi(\mathbf{f}_{V_1}, \{u_1, u_2, u_3, u_4\}), \\ &= [P(\mathbf{g}_K) || N(\mathbf{g}_K)], \end{aligned}$$

with all the $u_j \in \{1,2\}$,

$$\begin{aligned} P(\mathbf{g}_K) &= \{(1,1,1,u_1),(1,2,2,u_2),(2,1,2,u_3),(2,2,1,u_4)\}, \\ N(\mathbf{g}_K) &= \{(2,1,1,u_1),(2,2,2,u_2),(1,1,2,u_3),(1,2,1,u_4)\}, \end{aligned}$$

and the negatives of these moves. By symmetry, a Markov basis $\mathcal{H}_2$ for $\mathbf{T}(\{2,3\},\{2,4\},\{3,4\})$ consists of the move $\mathbf{f}_{V_2} = [P(\mathbf{f}_{V_2})||N(\mathbf{f}_{V_2})]$, where

$$\begin{aligned} P(\mathbf{f}_{V_2}) &= \{(1,1,1),(1,2,2),(2,1,2),(2,2,1)\}, \\ N(\mathbf{f}_{V_2}) &= \{(2,1,1),(2,2,2),(1,1,2),(1,2,1)\}, \end{aligned}$$

and its negative. The set of moves $\text{Ext}(\mathcal{H}_2 \to \mathbf{T}(D_1,\ldots,D_r))$ consists of the sixteen moves $\mathbf{g}_K = \psi(\mathbf{f}_{V_2},\{u_1,u_2,u_3,u_4\})$ with all $u_j \in \{1,2\}$,

$$\begin{aligned} P(\mathbf{g}_K) &= \{(u_1,1,1,1),(u_2,1,2,2),(u_3,2,1,2),(u_4,2,2,1)\}, \\ N(\mathbf{g}_K) &= \{(u_3,2,1,1),(u_4,2,2,2),(u_1,1,1,2),(u_1,1,2,1)\}, \end{aligned}$$

and the negatives of these moves. The set $\mathcal{F}(\{1,2,3\},\{2,3,4\})$ consists of the four moves of the form $\mathbf{f}_K = [P(\mathbf{f}_K)||N(\mathbf{f}_K)]$, where

$$\begin{aligned} P(\mathbf{f}_K) &= \{(1,u_1,u_2,1),(2,u_1,u_2,2)\}, \\ N(\mathbf{f}_K) &= \{(1,u_1,u_2,2),(2,u_1,u_2,1)\}, \end{aligned}$$

with all $u_j \in \{1,2\}$, and their negatives. Theorem 5.6 says that a Markov basis for the class of tables $\mathbf{T}(\{1,2\},\{1,3\},\{2,3\},\{2,4\},\{3,4\})$ is given by the union of these three sets of moves. ∎

For the purpose of increasing the efficiency of computation, it is desirable to have a Markov basis that is as small as possible. Sullivant (2002) shows how to construct a Markov basis smaller than the one described in Theorem 5.6. The main idea behind his construction is to note that the moves in $\mathcal{F}(V_1,V_2)$ can be used to manipulate moves as well as tables. It then turns out that some of the moves in $\text{Ext}(\mathcal{H}_1 \to \mathbf{T}(D_1,\ldots,D_r)) \cup \text{Ext}(\mathcal{H}_2 \to \mathbf{T}(D_1,\ldots,D_r))$ could be eliminated from $\mathcal{H}$ such that the remaining moves are still a Markov basis for $\mathbf{T}(D_1,\ldots,D_r)$.

6 Conclusions

Direct algebraic computations that work well for low-dimensional examples prove to be impractical for the high-dimensional problems that arise in practice. The results presented in this paper demonstrate that using the underlying graphical structure of the class of tables being studied can greatly reduce the amount of computations necessary for generating a Markov basis. We have shown how Markov bases for classes of tables $\mathbf{T}(D_1,D_2,\ldots,D_r)$ whose marginals have a reducible structure can be built from the Markov bases for the component structures in a relatively simple way. We exploit, in this context, the theory of *reducible* graphs as they were introduced by Tarjan (1985) and Leimer (1993). This idea is not entirely new and was used, among others, by Dobra and Fienberg (2000) to calculate the maximum and the minimum values of a cell entry in a table in $\mathbf{T}(\mathbf{n}_{D_1}, \mathbf{n}_{D_2}, \ldots, \mathbf{n}_{D_r})$.

Theorem 5.6 is the main theorem of this paper. It gives the basic construction from which the divide-and-conquer procedure we propose is built. Besides structurally characterizing a Markov basis for reducible sets of tables, this result has practical applications. Indeed, it is easier to compute the Markov basis for a class of tables using a smaller number of variables (using algebraic methods or otherwise), so our procedure reduces one long computation in a large number of variables to a few much shorter computations, each in a relatively smaller number of variables.

One should note that given a particular reducible structure, there are possibly many different Markov bases that could be constructed by the procedure outlined in this paper. Further research is needed to understand whether stochastic sampling from $\mathbf{T}(\mathbf{n}_{D_1}, \mathbf{n}_{D_2}, \ldots, \mathbf{n}_{D_r})$ based on different Markov bases will lead to the same inferences.

Acknowledgments

The work of Adrian Dobra was supported in part by the National Science Foundation under Grant EIA-9876619 to the National Institute of Statistical Sciences. Seth Sullivant was supported in part under a National Science Foundation Graduate Research Fellowship. The authors would like to thank Stephen E. Fienberg, Serkan Hoşten and Bernd Sturmfels for their comments that significantly improved this material.

References

Agresti, A. (1992), 'A survey of exact inference for contingency tables (with discussion)', *Statistical Science*.

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, M.I.T. Press, Cambridge, MA.

Blair, J. R. S. & Barry, P. (1993), An introduction to chordal graphs and clique trees, *in* IMA, ed., 'Graph Theory and Sparse Matrix Computation', Vol. 56, Springer-Verlag, New York, pp. 1–30.

Cox, D., Little, J. & O'Shea, D. (1992), *Ideals, Varieties and Algorithms*, Springer-Verlag, New York.

Dalenius, T. & Reiss, S. P. (1982), 'Data-swapping: a technique for disclosure control', *Journal of Statistical Planning and Inference* **6**, 73–85.

De Loera, J. & Sturmfels, B. (2001), Algebraic unimodular counting, Manuscript.

Diaconis, P. & Efron, B. (1985), 'Testing for independence in a two-way table: New interpretations of the chi-square statistic', *The Annals of Statistics* **13**, 845–874.

Diaconis, P. & Gangolli, A. (1995), Rectangular arrays with fixed margins, *in* 'Discrete Probability and Algorithms', Springer-Verlag, New York, pp. 15–41.

Diaconis, P. & Sturmfels, B. (1998), 'Algebraic algorithms for sampling from conditional distributions', *The Annals of Statistics* **26**, 363–397.

Dinwoodie, I. H. (1998), 'The Diaconis-Sturmfels algorithm and rules of succession', *Bernoulli* **4**, 401–410.

Dobra, A. (2001), Markov bases for decomposable graphical models, Manuscript.

Dobra, A. (2002), Statistical Tools for Disclosure Limitation in Multi-way Contingency Tables, PhD thesis, Department of Statistics, Carnegie Mellon University.

Dobra, A. & Fienberg, S. E. (2000), 'Bounds for cell entries in contingency tables given marginal totals and decomposable graphs', *Proceedings of the National Academy of Sciences* **97**, 11885–11892.

Fienberg, S. E., Makov, E. U. & Steele, R. J. (1998), 'Disclosure limitation using perturbation and related methods for categorical data', *Journal of Official Statistics* **14**, 485–511.

Fienberg, S. E., Makov, U. E., Meyer, M. M. & Steele, R. J. (2001), Computing the exact distribution for a multi-way contingency table conditional on its marginals totals, *in* A. Saleh, ed., 'Data Analysis from Statistical Foundations: A Festschrift in Honor of the 75th Birthday of D. A. S. Fraser', Nova Science Publishers, Huntington, NY, pp. 145–165.

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

Leimer, H. G. (1993), 'Optimal decomposition by clique separators', *Discrete Mathematics* **113**, 99–123.

Madigan, D. & York, J. (1995), 'Bayesian graphical models for discrete data', *International Statistical Review* **63**, 215–232.

Mehta, C. (1994), 'The exact analysis of contingency tables in medical research', *Statistical Methods in Medical Research.*

Mount, J. (1995), Application of Convex Sampling to Optimization and Contingency Table Generation/Counting, PhD thesis, Carnegie Mellon University.

Sullivant, S. (2002), Algebraic geometry and combinatorics of hierarchical models, Master's thesis, San Francisco State University.

Takken, A. (1999), Monte Carlo Goodness-of-Fit Tests for Discrete Data, PhD thesis, Stanford University.

Tarjan, R. E. (1985), 'Decomposition by clique separators', *Discrete Mathematics* **55**, 221–232.

Vlach, M. (1986), 'Conditions for the existence of solutions of the three-dimensional planar transportation problem', *Discrete Applied Mathematics* **13**, 61–78.

Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, John Wiley & Sons. New York.

Willenborg, L. & de Waal, T. (2000), *Elements of Statistical Disclosure Control*, Vol. 155, Lecture Notes in Statistics. Springer-Verlag, New York.

A Decomposable and Reducible Graphs

A graph $\mathcal{G}$ is a pair (K, E), where $K = \{1, 2, \ldots, k\}$ is a finite set of vertices and $E \subseteq K \times K$ is a set of edges linking the vertices. For any vertex set $A \subseteq K$, we define the edge set associated with it as

$$E(A) := \{(u, v) \in E | u, v \in A\}.$$

Let $\mathcal{G}(A) = (A, E(A))$ denote the subgraph of $\mathcal{G}$ induced by A. Two vertices $u, v \in K$ are *adjacent* if $(u, v) \in E$. A set of vertices of $\mathcal{G}$ is *independent* if no two of its elements are adjacent. An induced subgraph $\mathcal{G}(A)$ is *complete* if the vertices in A are pairwise adjacent in $\mathcal{G}$. We also say that A is *complete* in $\mathcal{G}$. A complete vertex set A in $\mathcal{G}$ that is maximal is a *clique*.

Let $u, v \in K$. A *path* (or *chain*) from u to v is a sequence $u = v_0, \ldots, v_n = v$ of distinct vertices such that $(v_{i-1}, v_i) \in E$ for all $i = 1, 2, \ldots, n$. The path is a *cycle* if the end points are allowed to be the same, $u = v$. If there is a path from u to v we say that u and v are *connected*. The sets $A, B \subset K$ are *disconnected* if u and v are not connected for all $u \in A$, $v \in B$. The *connected component* of a vertex $u \in K$ is the set of all vertices connected with u. A graph is *connected* if all the pairs of vertices are connected.

The set $C \subset K$ is an *uv-separator* if all paths from u to v intersect C. The set $C \subset K$ *separates* A from B if it is an uv-separator for every $u \in A$, $v \in B$. C is a *separator* of $\mathcal{G}$ if two vertices in the same connected component of $\mathcal{G}$ are in two distinct connected components of $\mathcal{G} \setminus C$ or, equivalently, if $\mathcal{G} \setminus C$ is disconnected. In addition, C is a *minimal* separator of $\mathcal{G}$ if C is a separator and no proper subset of C separates the graph. Unless otherwise stated, the separators we work with will be complete.

Decomposable graphs possess the special property that allows us to "decompose" them into components or subgraphs and work directly with these components. The idea is to decompose the graph $\mathcal{G}$ in two possibly overlapping subgraphs $\mathcal{G}'$ and $\mathcal{G}''$ so that no information of the graph is lost when transforming $\mathcal{G}$ into $\mathcal{G}'$ and $\mathcal{G}''$. Furthermore, by "correctly" decomposing $\mathcal{G}'$ and $\mathcal{G}''$, and so on, one ends up with a set of subgraphs of $\mathcal{G}$ which allow for no further decompositions. A set of subgraphs of $\mathcal{G}$ generated in this way is called a *derived system* of $\mathcal{G}$, while its elements are called *atoms* (Tarjan 1985). We define what we mean by "correct" decomposition.

Definition A.1. The partition (A_1, S, A_2) of K is said to form a *decomposition* of $\mathcal{G}$ if S is a minimal separator of A_1 and A_2.

In this case (A_1, S, A_2) *decomposes* $\mathcal{G}$ into the *components* $\mathcal{G}(A_1 \cup S)$ and $\mathcal{G}(S \cup A_2)$. The decomposition is *proper* if A_1 and A_2 are not empty.

Definition A.2. The graph $\mathcal{G}$ is *decomposable* if it is complete or if there exists a proper decomposition (A_1, S, A_2) into decomposable graphs $\mathcal{G}(A_1 \cup S)$ and $\mathcal{G}(S \cup A_2)$.

Graphs that are not decomposable, but can still be decomposed in sequences of atoms are described in Tarjan (1985) and Leimer (1993). In this case, the resulting atoms are not necessarily complete.

Definition A.3. A graph $\mathcal{G}$ is *reducible* if $\mathcal{G}$ admits a proper decomposition, otherwise $\mathcal{G}$ is a *prime* graph.

Given that every reducible graph $\mathcal{G}$ might have several derived systems (Tarjan 1985), we would like to be able to isolate one of them which could fully characterize the input graph $\mathcal{G}$.

Definition A.4. A subgraph $\mathcal{G}(A)$ is a *maximal prime (mp-) subgraph* of $\mathcal{G}$, if $\mathcal{G}(A)$ is prime and $\mathcal{G}(B)$ is reducible for all B with $A \subset B \subseteq K$.

The set of mp-subgraphs of $\mathcal{G}$ is contained in every derived system of $\mathcal{G}$. Moreover, the set of mp-subgraphs of $\mathcal{G}$ is always a derived system of $\mathcal{G}$ (Leimer 1993), and consequently it is the *unique minimal derived system.* If $\mathcal{G}$ is decomposable, the mp-subgraphs of $\mathcal{G}$ are complete, hence the unique minimal derived system of a decomposable graph contains only its cliques (Leimer 1993).