Some characterizations of minimal Markov basis for sampling from discrete conditional distributions

Akimichi TAKEMURA and Satoshi AOKI

Graduate School of Information Science and Technology
University of Tokyo, Tokyo, Japan

SUMMARY

In this paper we give some basic characterizations of minimal Markov basis for a connected Markov chain, which is used for performing exact tests in discrete exponential families given a sufficient statistic. We also give a necessary and sufficient condition for uniqueness of minimal Markov basis. A general algebraic algorithm for constructing a connected Markov chain was given by Diaconis and Sturmfels (1998). Their algorithm is based on computing Gröbner basis for a certain ideal in a polynomial ring, which can be carried out by using available computer algebra packages. However structure and interpretation of Gröbner basis produced by the packages are not necessarily clear, due to the lack of symmetry and minimality inherent in Gröbner basis computation. Our approach clarifies partially ordered structure of minimal Markov basis.

1 Introduction

In performing exact conditional tests in discrete exponential families given a sufficient statistic, the p values are usually calculated by large sample approximations. However when the data are sparse, i.e., when the sample size is small compared to the size of the sample space, large sample approximation may not be sufficiently accurate. When the sample size and the sample space are small, enumeration of the sample space may be feasible with some ingenious enumeration schemes. For the case of two-way contingency tables with fixed row and column sums, Mehta and Patel (1983) proposed a network algorithm, which incorporates appropriate trimming in enumeration. Aoki (2001, 2002) proposed improvement of this trimming for two-way contingency tables and a conditional test of Hardy-Weinberg model.

For sparse data sets, where enumeration becomes infeasible and the large sample approximation is not adequate, approximation of the p values by Monte Carlo methods may be the only feasible approach. For some models, random sample can be easily generated from the conditional distribution. A primary example is decomposable log-linear models (e.g., Section 4.4 of Lauritzen (1996)) in multi-way contingency tables. In decomposable models, random sample can be easily generated by exploiting the nesting structure of conditional independence. For testing more general hierarchical log-linear models in multi-way contingency tables, direct generation of random sample is difficult. In this case Markov chain Monte Carlo techniques can be used.

As an example, consider testing the null hypothesis of no three-factor interactions in the log-linear model of three-way contingency tables. This is the simplest case of non-decomposable hierarchical log-linear model for multi-way contingency tables and we want to sample from the hypergeometric distribution over three-way tables with all two-dimensional marginal frequencies fixed. This problem is surprisingly difficult. Not only the direct generation of random samples but also the construction of appropriate connected Markov chain is difficult as shown in Diaconis and Sturmfels (1998). For the special case of $3 \times 3 \times K$ tables, we gave an explicit form of unique minimal basis for connected Markov chain (Aoki and Takemura (2002a)). After laborious derivation we found that the minimal Markov basis for the $3 \times 3 \times K$ case consists of moves of degrees 4, 6, 8 and 10 (see Section 3 below). At present it seems very difficult to obtain explicit forms of minimal Markov basis for general three-way tables. It should be even harder to completely characterize minimal Markov basis for testing general hierarchical log-linear models in m-way contingency tables.

For the case of two-way tables, it is rather simple to describe connected Markov chain over two-way tables with fixed one-dimensional marginals, if there are no additional restrictions on individual cells. However if there are additional restrictions on the cell frequencies such as structural zeros in two-way tables, description of a connected Markov chain becomes more complicated. In our subsequent work (Aoki and Takemura (2002b)) we give an explicit characterization of unique minimal Markov basis for two-way tables with fixed marginals and arbitrary pattern of structural zeros. See Chapter 5 of Bishop et al. (1975) for comprehensive treatment of structural zeros in contingency tables.

In principle the Gröbner basis technique proposed in Diaconis and Sturmfels (1998) can be employed to obtain a basis for connected Markov chain over any given sample space. See also Dinwoodie (1998) for a clear exposition of the Gröbner basis technique. However the Gröbner basis computation is very time consuming for large sample space. Furthermore Gröbner basis computation produces large number of redundant basis elements due to the lack of symmetry and minimality inherent in Gröbner basis. The lack of symmetry stems from the dependence of Gröbner basis on the particular term order. For example, Sakata and Sawae (2000) reports that for $4 \times 4 \times 4$ tables with fixed two-dimensional marginals, their lattice based algorithm was not able to produce the Gröbner basis after two months of computation and the incomplete basis at that time already contained more than 340,000 basis elements. Furthermore the resulting Gröbner basis may not be easily interpretable due to the redundant elements and the dependence on the chosen term order. Therefore at present Gröbner basis technique does not seem to be a satisfactory answer.

In this paper we derive some basic characterizations of minimal Markov basis for a connected Markov chain for sampling from conditional distributions. Our arguments are totally elementary. We also give a necessary and sufficient condition for the uniqueness of minimal Markov basis. Our approach is basically constructive and it clarifies a partially ordered structure of minimal Markov basis. At present our result is not powerful enough to completely characterize minimal Markov basis for a given problem, but with further refinement it might be possible to implement an alternative algorithm for constructing Markov basis for a connected Markov chain over a given sample space.

In Section 2 we derive our characterization of minimal Markov basis. Relevant examples of discrete exponential families are studied in Section 3.

2 Characterization and construction of minimal Markov basis

In this section we first give necessary notations and definitions. We follow the framework of Diaconis and Sturmfels (1998). Our notation is somewhat different from theirs. Then we state our results on minimal Markov basis and its uniqueness. In Section 2.3 we give proofs of our results and some additional facts.

2.1 Notations and definitions

Let $\mathcal{I}$ be a finite set. With contingency tables in mind, an element of $\mathcal{I}$ is called a *cell* and denoted by $\boldsymbol{i} \in \mathcal{I}$. $|\mathcal{I}|$ denotes the number of cells. In the case of $I_1 \times \cdots \times I_m$ m-way contingency tables, $\boldsymbol{i}$ represents a multi-index $\boldsymbol{i} = (i_1, \ldots, i_m)$ and $|\mathcal{I}| = I_1 \times \cdots \times I_m$. A non-negative integer $x_{\boldsymbol{i}}$ denotes the frequency of cell $\boldsymbol{i}$. $n = \sum_{\boldsymbol{i} \in \mathcal{I}} x_{\boldsymbol{i}}$ denotes the sample size.

Let $\ = \{0, 1, \ldots\}$ denote the set of non-negative integers and let $\ = \{0, \pm 1, \ldots\}$ denote the set of integers. Let $a(\boldsymbol{i}) \in {}^{\nu}$, $\boldsymbol{i} \in \mathcal{I}$, denote ν-dimensional fixed column vectors consisting of non-negative integers. A ν-dimensional sufficient statistic $\boldsymbol{t}$ is given by

$$\boldsymbol{t} = \sum_{\boldsymbol{i} \in \mathcal{I}} a(\boldsymbol{i}) x_{\boldsymbol{i}}.$$

In the case of hierarchical model for m-way contingency tables $\boldsymbol{t}$ consists of appropriate marginal totals.

Let the cells and the vectors $a(\boldsymbol{i})$ be appropriately ordered. For m-way contingency tables, we may order the multi-indices lexicographically. Let

$$\boldsymbol{x} = \{x_{\boldsymbol{i}}\}_{\boldsymbol{i} \in \mathcal{I}} \in {}^{|\mathcal{I}|}$$

denote an $|\mathcal{I}|$-dimensional column vector of cell frequencies and let

$$A = \{a(\boldsymbol{i})\}_{\boldsymbol{i} \in \mathcal{I}}$$

denote a $\nu \times |\mathcal{I}|$ matrix. Then the sufficient statistic $\boldsymbol{t}$ is written as

$$\boldsymbol{t} = A\boldsymbol{x}.$$

We call $\boldsymbol{x}$ a *frequency vector*. We also use the notation $|\boldsymbol{x}| = n = \sum_{\boldsymbol{i}} x_{\boldsymbol{i}}$ to denote the sample size and the notation $\boldsymbol{x} \geq 0$ to denote that the elements of $\boldsymbol{x}$ are non-negative integers. The set of $\boldsymbol{x}$'s for a given $\boldsymbol{t}$ is denoted by

$$\mathcal{F}_{\boldsymbol{t}} = \{\boldsymbol{x} \geq 0 \mid A\boldsymbol{x} = \boldsymbol{t}\}.$$

Concerning the matrix A we make the following assumption throughout this paper.

Assumption 1 *$|\mathcal{I}|$-dimensional row vector $(1, 1, \ldots, 1)$ is a linear combination of the rows of A.*

Assumption 1 is satisfied in the examples of Section 3. Assumption 1 implies that the sample size n is determined from the sufficient statistic $\boldsymbol{t}$ and all elements of $\mathcal{F}_{\boldsymbol{t}}$ have the same sample size. Somewhat abusing the notation, we write $n = |\boldsymbol{t}|$ to denote the sample size of elements of $\mathcal{F}_{\boldsymbol{t}}$. Another consequence of this assumption is that each $a(\boldsymbol{i})$, $\boldsymbol{i} \in \mathcal{I}$, is a non-zero vector, because otherwise each linear combination of the rows of A has 0 in the $\boldsymbol{i}$-th position.

For the case of $I \times J$ contingency tables with fixed one-dimensional marginals and with lexicographical ordering of cells, A is written as

$$A = \begin{bmatrix} 1_I' \otimes E_J \\ E_I \otimes 1_J' \end{bmatrix},$$

where 1_I is the I-dimensional vector consisting of 1's, E_J is the $J \times J$ identity matrix and $\otimes$ denotes the Kronecker product. Similarly for $I \times J \times K$ contingency tables with fixed two-dimensional marginals and with lexicographical ordering of cells, A is written as

$$A = \begin{bmatrix} 1_I' \otimes E_J \otimes E_K \\ E_I \otimes 1_J' \otimes E_K \\ E_I \otimes E_J \otimes 1_K' \end{bmatrix}. \tag{1}$$

An $|\mathcal{I}|$-dimensional vector of integers $\boldsymbol{z} \in {}^{|\mathcal{I}|}$ is called a *move* if it is in the kernel of A:

$$A\boldsymbol{z} = 0.$$

Adding a move $\boldsymbol{z}$ to $\boldsymbol{x}$ does not change the sufficient statistic

$$\boldsymbol{t} = A\boldsymbol{x} = A(\boldsymbol{x} + \boldsymbol{z}).$$

Therefore $\boldsymbol{z}$ can be interpreted as a move within $\mathcal{F}_{\boldsymbol{t}}$ for any $\boldsymbol{t}$. By definition the zero frequency vector $\boldsymbol{z} = 0$ is also a move, although it does not move anything. For a move $\boldsymbol{z}$, the positive part $\boldsymbol{z}^+$ and the negative part $\boldsymbol{z}^-$ are defined by

$$z_{\boldsymbol{i}}^+ = \max(z_{\boldsymbol{i}}, 0), \qquad z_{\boldsymbol{i}}^- = -\min(z_{\boldsymbol{i}}, 0),$$

respectively. Then $\boldsymbol{z} = \boldsymbol{z}^+ - \boldsymbol{z}^-$. Note that if $\boldsymbol{z}$ is a move, then $-\boldsymbol{z}$ is also a move with $(-\boldsymbol{z})^+ = \boldsymbol{z}^-$ and $(-\boldsymbol{z})^- = \boldsymbol{z}^+$. Note also that non-zero elements of $\boldsymbol{z}^+$ and $\boldsymbol{z}^-$ do not share a common cell. The positive part $\boldsymbol{z}^+$ and the negative part $\boldsymbol{z}^-$ have the same value of sufficient statistic $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$. The sample size of $\boldsymbol{z}^+$ (or $\boldsymbol{z}^-$) is called the *degree* of $\boldsymbol{z}$ and denoted by

$$\deg \boldsymbol{z} = |\boldsymbol{z}^+| = |\boldsymbol{z}^-|.$$

Occasionally we also write $|\boldsymbol{z}| = \sum_{\boldsymbol{i}} |z_{\boldsymbol{i}}| = 2 \deg \boldsymbol{z}$.

We say that a move $\boldsymbol{z}$ is *applicable* to $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{t}}$ if $\boldsymbol{x} + \boldsymbol{z} \in \mathcal{F}_{\boldsymbol{t}}$, i.e., adding $\boldsymbol{z}$ to $\boldsymbol{x}$ does not produce a negative cell. Since

$$\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{x} + \boldsymbol{z}^+ - \boldsymbol{z}^-,$$

$\boldsymbol{z}$ is applicable to $\boldsymbol{x}$ if and only if

$$\boldsymbol{x} \geq \boldsymbol{z}^-. \tag{2}$$

Note also that $\boldsymbol{z}$ is applicable to $\boldsymbol{x}$ if and only if $-\boldsymbol{z}$ is applicable to $\boldsymbol{x} + \boldsymbol{z}$.

Let $\mathcal{B} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L\}$ be a finite set of moves. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}_{\boldsymbol{t}}$. We say that $\boldsymbol{y}$ is *accessible* from $\boldsymbol{x}$ by $\mathcal{B}$ and denote it by

$$\boldsymbol{x} \sim \boldsymbol{y} \qquad (\text{mod } \mathcal{B}),$$

if there exists a sequence of moves $\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_k}$ from $\mathcal{B}$ and $\epsilon_j = \pm 1$, $j = 1, \ldots, k$, such that $\boldsymbol{y} = \boldsymbol{x} + \sum_{j=1}^{k} \epsilon_j \boldsymbol{z}_{i_j}$ and

$$\boldsymbol{x} + \sum_{j=1}^{h} \epsilon_j \boldsymbol{z}_{i_j} \in \mathcal{F}_{\boldsymbol{t}}, \qquad h = 1, \ldots, k-1, \tag{3}$$

i.e., we can move from $\boldsymbol{x}$ to $\boldsymbol{y}$ by moves from $\mathcal{B}$ without causing negative cells on the way. Obviously the notion of accessibility is symmetric and transitive:

$$\begin{aligned} &\boldsymbol{x} \sim \boldsymbol{y} \quad \Rightarrow \quad \boldsymbol{y} \sim \boldsymbol{x} \qquad (\text{mod } \mathcal{B}), \\ &\boldsymbol{x}_1 \sim \boldsymbol{x}_2, \quad \boldsymbol{x}_2 \sim \boldsymbol{x}_3 \quad \Rightarrow \quad \boldsymbol{x}_1 \sim \boldsymbol{x}_3 \qquad (\text{mod } \mathcal{B}). \end{aligned}$$

Therefore accessibility by $\mathcal{B}$ is an equivalence relation and each $\mathcal{F}_{\boldsymbol{t}}$ is partitioned into disjoint equivalence classes by moves of $\mathcal{B}$. We call these equivalence classes $\mathcal{B}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$. Since the notion of accessibility symmetric, we also say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are mutually accessible by $\mathcal{B}$ if $\boldsymbol{x} \sim \boldsymbol{y}$ (mod $\mathcal{B}$). Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be elements from two different $\mathcal{B}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$. We say that a move

$$\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$$

connects these two equivalence classes. Diaconis, Eisenbud and Sturmfels (1998) gives results on properties of a $\mathcal{B}$-equivalence class from algebraic viewpoint.

Particular sets of moves we consider below are

$$\mathcal{B}_{\boldsymbol{t}} = \{\boldsymbol{z} \mid \boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-\},$$

which is a set of moves $\boldsymbol{z}$ with the same value of the sufficient statistic $\boldsymbol{t} = A\boldsymbol{z}^+$, and

$$\mathcal{B}_n = \{\boldsymbol{z} \mid \deg \boldsymbol{z} \le n\},$$

which is a set of moves with degree less than or equal to n.

A set of finite moves $\mathcal{B} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L\}$ is a *Markov basis* if for all $\boldsymbol{t}$, $\mathcal{F}_{\boldsymbol{t}}$ itself constitutes one $\mathcal{B}$-equivalence class, i.e., for every $\boldsymbol{t}$ and for every $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}_{\boldsymbol{t}}$, $\boldsymbol{y}$ and $\boldsymbol{x}$ are mutually accessible by $\mathcal{B}$. Logically important point here is the existence of a finite Markov basis, which is guaranteed by the Hilbert basis theorem (see Section 3.1 of Diaconis and Sturmfels (1998)). In fact Diaconis and Sturmfels (1998) gave an algorithm to produce a finite Markov basis, although in practice their algorithm does not seem to be satisfactory at present as already discussed in Section 1. A Markov basis $\mathcal{B}$ is *minimal* if no proper subset of $\mathcal{B}$ is a Markov basis. A minimal Markov basis always exists, because from any Markov basis, we can remove redundant elements one by one, until none of the remaining elements can be removed any further.

It should be noted that a Markov basis $\mathcal{B}$ is common for all $\boldsymbol{t}$. Suppose that a data frequency vector $\boldsymbol{x}$ is given and we are concerned only with connecting frequency vectors of $\mathcal{F}_{\boldsymbol{t}}$ for the given $\boldsymbol{t} = A\boldsymbol{x}$. Then we may not need all of the moves from $\mathcal{B}$. It is a subtle problem to determine which moves of $\mathcal{B}$ are needed for connecting $\mathcal{F}_{\boldsymbol{t}}$ for a given $\boldsymbol{t}$ from the viewpoint of minimality. We discuss this point further in Section 3.2. We also investigate this problem for the case of two-way contingency tables with structural zeros in our subsequent work (Aoki and Takemura (2002b)).

Having prepared adequate notations and definitions, we now proceed to characterize structure of the minimal Markov basis.

2.2 Main results

For each $\boldsymbol{t}$, let $n = |\boldsymbol{t}|$ be the sample size of elements of $\mathcal{F}_{\boldsymbol{t}}$ and let $\mathcal{B}_{n-1}$ be the set of moves with degree less than n. Write the $\mathcal{B}_{n-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$ as

$$\mathcal{F}_{\boldsymbol{t}} = \mathcal{F}_{\boldsymbol{t},1} \cup \cdots \cup \mathcal{F}_{\boldsymbol{t},K_{\boldsymbol{t}}}. \tag{4}$$

Let $\boldsymbol{x}_j \in \mathcal{F}_{\boldsymbol{t},j},\ j = 1, \ldots, K_{\boldsymbol{t}}$, be representative elements of the equivalence classes and

$$\boldsymbol{z}_{j_1,j_2} = \boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_2}, \quad j_1 \neq j_2$$

be a move connecting $\mathcal{F}_{\boldsymbol{t},j_1}$ and $\mathcal{F}_{\boldsymbol{t},j_2}$. Note that we can connect all equivalence classes with $K_{\boldsymbol{t}} - 1$ moves of this type, by forming a tree, where the equivalence classes are interpreted as vertices and connecting moves are interpreted as edges of a graph. Now we state our main theorem.

Theorem 1 *Let $\mathcal{B}$ be a minimal Markov basis. For each $\boldsymbol{t}$, $\mathcal{B} \cap \mathcal{B}_{\boldsymbol{t}}$ consists of $K_{\boldsymbol{t}} - 1$ moves connecting different $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$, in such a way that the equivalence classes are connected into a tree by these moves.*

Conversely choose any $K_{\boldsymbol{t}}-1$ moves $z_{\boldsymbol{t},1}, \ldots, z_{\boldsymbol{t},K_{\boldsymbol{t}}-1}$ connecting different $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$, in such a way that the equivalence classes are connected into a tree by these moves. Then

$$\mathcal{B} = \bigcup_{\boldsymbol{t}:K_{\boldsymbol{t}} \geq 2} \{z_{\boldsymbol{t},1}, \ldots, z_{\boldsymbol{t},K_{\boldsymbol{t}}-1}\} \tag{5}$$

is a minimal Markov basis.

Note that no move is needed from $\mathcal{F}_{\boldsymbol{t}}$ with $K_{\boldsymbol{t}} = 1$, including the case where $\mathcal{F}_{\boldsymbol{t}}$ is a one-element set. If $\mathcal{F}_{\boldsymbol{t}} = \{\boldsymbol{x}\}$ is a one-element set, no non-zero move is applicable to $\boldsymbol{x}$, but at the same time we do not need to move from $\boldsymbol{x}$ at all for such a $\boldsymbol{t}$.

In principle this theorem can be used to construct a minimal Markov basis from below as follows. As the initial step we consider $\boldsymbol{t}$ with the sample size $n = |\boldsymbol{t}| = 1$. Because $\mathcal{B}_0$ consists only of the zero move $\mathcal{B}_0 = \{0\}$, each point $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{t}},\ |\boldsymbol{t}| = 1$, is isolated and forms an equivalence class by itself. For each $\boldsymbol{t}$ with $|\boldsymbol{t}| = 1$, we choose $K_{\boldsymbol{t}} - 1$ degree 1 moves to connect $K_{\boldsymbol{t}}$ points of $\mathcal{F}_{\boldsymbol{t}}$ into a tree. Let $\tilde{\mathcal{B}}_1$ be the set of chosen moves. $\tilde{\mathcal{B}}_1$ is a subset of the set $\mathcal{B}_1$ of all degree 1 moves. Since every degree 1 move can be expressed by non-negative integer combination of chosen degree 1 moves, it follows that $\tilde{\mathcal{B}}_1$ and $\mathcal{B}_1$ induce same equivalence classes for each $\mathcal{F}_{\boldsymbol{t}}$ with $|\boldsymbol{t}| = 2$. Therefore as the second step we consider $\tilde{\mathcal{B}}_1$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$ for each $\boldsymbol{t}$ with $|\boldsymbol{t}| = 2$ and choose representative elements from each equivalence class to form degree 2 moves connecting the equivalence classes into a tree. We add the chosen moves to $\tilde{\mathcal{B}}_1$ and form a set $\tilde{\mathcal{B}}_2$. We can repeat this process for $n = |\boldsymbol{t}| = 3, 4, \ldots$. By the Hilbert basis theorem there exists some n_0 such that for $n \geq n_0$ no new moves need to be added. Then a minimal Markov basis $\mathcal{B}$ of (5) is written as $\mathcal{B} = \tilde{\mathcal{B}}_{n_0}$. Obviously there is a considerable difficulty in implementing this procedure. We will discuss this point further in Section 3.2.

Theorem 1 clarifies to what extent minimal Markov basis is unique. If an equivalence class consists of more than one element, then any element can be chosen as the representative element of the equivalence class. Another indeterminacy is how to form a tree of the equivalence classes. Considering these indeterminacies and in view of Lemma 3 below, we have the following corollary to Theorem 1.

Corollary 1 *Minimal Markov basis is unique if and only if for each* $\boldsymbol{t}$, $\mathcal{F}_{\boldsymbol{t}}$ *itself constitutes one* $\mathcal{B}_{|\boldsymbol{t}|-1}$*-equivalence class or* $\mathcal{F}_{\boldsymbol{t}}$ *is a two element set.*

In this corollary, the two cases are not necessarily exclusive, namely, there are cases where $\mathcal{F}_{\boldsymbol{t}}$ is a two element set forming a single $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence class. From this corollary it seems that minimal Markov basis is unique only under special conditions. It is therefore of great interest that minimal Markov basis is unique for some standard problems in m-way ($m \geq 2$) contingency tables with fixed marginals. On the other hand for the simplest case of one-way contingency tables, minimal Markov basis is not unique. These facts will be confirmed in Section 3

2.3 Proofs and some additional facts

Here we give a proof of Theorem 1. In the course of proving Theorem 1, we derive several lemmas, which may be of some independent interest.

Lemma 1 *If a move* $\boldsymbol{z}$ *is applicable to at least one element of* $\mathcal{F}_{\boldsymbol{t}}$*, then*

$$\deg \boldsymbol{z} \leq |\boldsymbol{t}|, \tag{6}$$

where the equality holds if and only if $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$.

Proof. Let $\boldsymbol{z}$ be applicable to $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{t}}$. Then by (2) $x_{\boldsymbol{i}} \geq z^-_{\boldsymbol{i}}$, $\forall \boldsymbol{i} \in \mathcal{I}$. Summing over $\mathcal{I}$ yields (6).

Concerning the equality, if $\boldsymbol{z}$ be applicable to $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{t}}$ and the equality holds in (6), then $x_{\boldsymbol{i}} = z^-_{\boldsymbol{i}}$, $\forall \boldsymbol{i} \in \mathcal{I}$ and

$$\boldsymbol{t} = A\boldsymbol{x} = \sum_{\boldsymbol{i} \in \mathcal{I}} a(\boldsymbol{i}) x_{\boldsymbol{i}} = \sum_{\boldsymbol{i} \in \mathcal{I}} a(\boldsymbol{i}) z^-_{\boldsymbol{i}} = A\boldsymbol{z}^-.$$

Conversely if $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$, then $\deg \boldsymbol{z} = |\boldsymbol{t}|$ by definition of $\deg \boldsymbol{z}$ and $|\boldsymbol{t}|$. Q.E.D.

Lemma 1 implies that in considering mutual accessibility between $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}_{\boldsymbol{t}}$, we only need to consider moves of degree smaller than $|\boldsymbol{t}|$ or moves $\boldsymbol{z}$ with $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$. Lemma 1 also implies the following simple but useful fact.

Lemma 2 *Suppose that* $\mathcal{F}_{\boldsymbol{t}} = \{\boldsymbol{x}, \boldsymbol{y}\}$ *is a two-element set and positive elements of* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *do not share a common cell. Then* $K_{\boldsymbol{t}} = 2$ *and* $\boldsymbol{x}, \boldsymbol{y}$ *are* $\mathcal{B}_{|\boldsymbol{t}|-1}$*-equivalence classes by themselves. Furthermore* $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{x}$ *belongs to each Markov basis.*

Proof. Suppose that $\boldsymbol{y}$ is accessible from $\boldsymbol{x}$ by $\mathcal{B}_{|\boldsymbol{t}|-1}$. Then there exists a non-zero move $\boldsymbol{z}$ with $\deg \boldsymbol{z} \le |\boldsymbol{t}|-1$ such that $\boldsymbol{z}$ is applicable to $\boldsymbol{x}$. If $\boldsymbol{x}+\boldsymbol{z}=\boldsymbol{y}$, then $\boldsymbol{z}=\boldsymbol{y}-\boldsymbol{x}$ and $\deg \boldsymbol{z} = |\boldsymbol{t}|$ because positive elements of $\boldsymbol{x}$ and $\boldsymbol{y}$ do not share a common cell. Therefore $\boldsymbol{x}+\boldsymbol{z} \neq \boldsymbol{y}$ and $\mathcal{F}_{\boldsymbol{t}}$ contains a third element $\boldsymbol{x}+\boldsymbol{z}$, which is a contradiction. Therefore $\boldsymbol{y}$ and $\boldsymbol{x}$ are in different $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes, implying that $\boldsymbol{y}$ and $\boldsymbol{x}$ are $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes by themselves.

Now consider moving from $\boldsymbol{x}$ to $\boldsymbol{y}$. Since they are $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes by themselves, no non-zero move $\boldsymbol{z}$ of degree $\deg \boldsymbol{z} < |\boldsymbol{t}|$ is applicable to $\boldsymbol{x}$. By Lemma 1, only moves $\boldsymbol{z}$ with $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$ are applicable to $\boldsymbol{x}$. If any such move is different from $\boldsymbol{y}-\boldsymbol{x}$, then as above $\mathcal{F}_{\boldsymbol{t}}$ contains a third element. It follows that in order to move from $\boldsymbol{x}$ to $\boldsymbol{y}$, we have to move by exactly one step using the move $\boldsymbol{z}=\boldsymbol{y}-\boldsymbol{x}$. Therefore $\boldsymbol{z}$ has to belong to any Markov basis. Q.E.D.

Define $\min(\boldsymbol{x},\boldsymbol{y})$, the minimum of $\boldsymbol{x}$ and $\boldsymbol{y}$, elementwise

$$\min(\boldsymbol{x},\boldsymbol{y})_{\boldsymbol{i}} = \min(x_{\boldsymbol{i}}, y_{\boldsymbol{i}}). \tag{7}$$

Lemma 2 can be slightly modified to yield the following result.

Lemma 3 *Suppose that $\mathcal{F}_{\boldsymbol{t}} = \{\boldsymbol{x},\boldsymbol{y}\}$ is a two-element set. Then $\boldsymbol{z}=\boldsymbol{y}-\boldsymbol{x}$ belongs to each Markov basis.*

Proof. If positive elements of $\boldsymbol{x}$ and $\boldsymbol{y}$ do not share a common cell, then the result is already contained in Lemma 2. Otherwise let $\boldsymbol{v}=\min(\boldsymbol{x},\boldsymbol{y})$ and consider $\boldsymbol{y}-\boldsymbol{v}$ and $\boldsymbol{x}-\boldsymbol{v}$. Non-zero elements of $\boldsymbol{y}-\boldsymbol{v}$ and $\boldsymbol{x}-\boldsymbol{v}$ do not share a common cell and by Lemma 2 again

$$\boldsymbol{z} = (\boldsymbol{y}-\boldsymbol{v}) - (\boldsymbol{x}-\boldsymbol{v}) = \boldsymbol{y}-\boldsymbol{x}$$

belongs to each Markov basis. Q.E.D.

For a frequency vector $\boldsymbol{x} = \{x_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{I}}$, define its *support* $\mathrm{supp}(\boldsymbol{x})$ by

$$\mathrm{supp}(\boldsymbol{x}) = \{\boldsymbol{i} \mid x_{\boldsymbol{i}} > 0\},$$

which is the set of positive cells of $\boldsymbol{x}$. For some set $\mathcal{F}$ of frequency vectors, define its support by

$$\mathrm{supp}(\mathcal{F}) = \bigcup_{\boldsymbol{x}\in\mathcal{F}} \mathrm{supp}(\boldsymbol{x}) = \{\boldsymbol{i} \mid x_{\boldsymbol{i}} > 0 \text{ for some } \boldsymbol{x}\in\mathcal{F}\}.$$

Then we have the following lemma.

Lemma 4 *Consider the $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of (4). The supports of the equivalence classes* $\mathrm{supp}(\mathcal{F}_{\boldsymbol{t},1}),\dots,\mathrm{supp}(\mathcal{F}_{\boldsymbol{t},K_{\boldsymbol{t}}})$ *are disjoint.*

Proof. Suppose that there exist $\boldsymbol{x}\in\mathcal{F}_{\boldsymbol{t},j_1}$, $\boldsymbol{y}\in\mathcal{F}_{\boldsymbol{t},j_2}$, $j_1\neq j_2$, such that positive elements of $\boldsymbol{x}$ and $\boldsymbol{y}$ share a common cell. Let $\boldsymbol{v}=\min(\boldsymbol{x},\boldsymbol{y})$ and consider $\boldsymbol{y}-\boldsymbol{v}$ and $\boldsymbol{x}-\boldsymbol{v}$. Because $\boldsymbol{v}$ is a non-zero vector, the sample size becomes smaller

$$|\boldsymbol{x}-\boldsymbol{v}| = |\boldsymbol{y}-\boldsymbol{v}| < n = |\boldsymbol{x}| = |\boldsymbol{y}|.$$

Then

$$\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{x} = (\boldsymbol{y} - \boldsymbol{v}) - (\boldsymbol{x} - \boldsymbol{v})$$

has degree $\deg \boldsymbol{z} = |\boldsymbol{x} - \boldsymbol{v}| < n$. Now $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$ is accessible from $\boldsymbol{x}$ by a single move $\boldsymbol{z}$. This is a contradiction, because $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to different $\mathcal{B}_{n-1}$-equivalence classes. Q.E.D.

The final lemma concerns replacing a move by series of moves.

Lemma 5 *Let $\mathcal{B}$ be a set of moves and let $\boldsymbol{z}_0 \notin \mathcal{B}$ be another non-zero move. Assume that $\boldsymbol{z}_0^+$ is accessible from $\boldsymbol{z}_0^-$ by $\mathcal{B}$. Then for each $\boldsymbol{x}$, to which $\boldsymbol{z}_0$ is applicable, $\boldsymbol{x} + \boldsymbol{z}_0$ is accessible from $\boldsymbol{x}$ by $\mathcal{B}$.*

This lemma shows that if $\boldsymbol{z}_0^+$ is accessible from $\boldsymbol{z}_0^-$ by $\mathcal{B}$, then we can always replace $\boldsymbol{z}_0$ by a series of moves from $\mathcal{B}$.

Proof. Suppose that $\boldsymbol{z}_0$ is applicable to $\boldsymbol{x}$. Then $\boldsymbol{x} - \boldsymbol{z}_0^- \geq 0$ by (2). By the definition of accessibility (cf. (3)), we can move from $\boldsymbol{z}_0^-$ to $\boldsymbol{z}_0^+$ by moves from $\mathcal{B}$ without causing negative cells on the way. Then the same sequence of moves can be applied to $\boldsymbol{x}$ without causing negative cells on the way, leading from $\boldsymbol{x}$ to $\boldsymbol{x} + \boldsymbol{z}_0$. Q.E.D.

Now we are ready to prove Theorem 1 and Corollary 1.

Proof of Theorem 1. Let $\mathcal{B}$ be a minimal Markov basis. Note that $\mathcal{B}_n$ and $\mathcal{B} \cap \mathcal{B}_n$ induces the same equivalence classes in $\mathcal{F}_{\boldsymbol{t}}$, $|\boldsymbol{t}| = n + 1$. Fix a particular $\boldsymbol{t}$. Write

$$\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L\} = \mathcal{B} \cap \mathcal{B}_{\boldsymbol{t}}.$$

For any $j = 1, \ldots, L$, let

$$\boldsymbol{x} = \boldsymbol{z}_j^+, \ \boldsymbol{y} = \boldsymbol{z}_j^-.$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the same $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence class, then by Lemma 5, $\boldsymbol{z}_j$ can be replaced by a series of moves of lower degree from $\mathcal{B}$ and $\mathcal{B} \setminus \boldsymbol{z}_j$ remains to be a Markov basis. This contradicts the minimality of $\mathcal{B}$. Therefore $\boldsymbol{z}_j^+$ and $\boldsymbol{z}_j^-$ are in two different $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes connecting them. Now we consider a graph, whose vertices are $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$ and whose edges are moves $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L$. If this graph is not connected, there exist two frequency vectors $\boldsymbol{x}, \boldsymbol{y}$ in two different equivalence classes, which are not mutually accessible by $\mathcal{B}$. In view of Lemma 1, this is a contradiction because $\mathcal{B}$ is a Markov basis. On the other hand if the graph contains a cycle, then there exits $\boldsymbol{z}_j$, such that $\boldsymbol{z}_j^+$ and $\boldsymbol{z}_j^-$ are mutually accessible by $\mathcal{B} \setminus \boldsymbol{z}_j$. By Lemma 5 again, this contradicts the minimality of $\mathcal{B}$. It follows that the graph is a tree. Since any tree with $K_{\boldsymbol{t}}$ vertices has $K_{\boldsymbol{t}} - 1$ edges, $L = K_{\boldsymbol{t}} - 1$.

Reversing the above argument, it is now easy to see that if $K_{\boldsymbol{t}} - 1$ moves $z_{\boldsymbol{t},1}, \ldots, z_{\boldsymbol{t},K_{\boldsymbol{t}}-1}$ connecting different $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$ are chosen in such a way that the equivalence classes are connected into a tree by these moves, then

$$\mathcal{B} = \bigcup_{\boldsymbol{t}: K_{\boldsymbol{t}} \geq 2} \{z_{\boldsymbol{t},1}, \ldots, z_{\boldsymbol{t},K_{\boldsymbol{t}}-1}\}$$

is a minimal Markov basis. Q.E.D.

Proof of Corollary 1. From our argument preceding Corollary 1, it follows that if minimal Markov basis is unique then for each $\boldsymbol{t}$, $\mathcal{F}_{\boldsymbol{t}}$ itself constitutes one $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence class or $\mathcal{F}_{\boldsymbol{t}}$

is a two element set $\{\boldsymbol{x}_{\boldsymbol{t},1}, \boldsymbol{x}_{\boldsymbol{t},2}\}$, such that $\boldsymbol{x}_{\boldsymbol{t},1} \not\sim \boldsymbol{x}_{\boldsymbol{t},2} \pmod{\mathcal{B}_{|\boldsymbol{t}|-1}}$. Therefore we only need to prove the converse. Suppose that for each $\boldsymbol{t}$, $\mathcal{F}_{\boldsymbol{t}}$ itself constitutes one $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence class or $\mathcal{F}_{\boldsymbol{t}}$ is a two element set. By Lemma 3, for each two-element set $\mathcal{F}_{\boldsymbol{t}} = \{\boldsymbol{x}, \boldsymbol{y}\}$ the move $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{x}$ belongs to each Markov basis. However by Theorem 1 each minimal Markov basis consists only of these moves. Therefore minimal Markov basis is unique. Q.E.D.

3 Examples and some discussion

In this section we verify Theorem 1 for various problems. First we investigate standard contingency tables with fixed marginals. In addition to minimality results on various problems, our result on three-way contingency tables with fixed one-dimensional marginals appears to be new. Then we investigate some other models including a simple case of Poisson regression model and the Hardy-Weinberg model. Finally we give some discussions.

3.1 Examples

One-way contingency tables. We start with the simplest case of one-way contingency tables. Let $\boldsymbol{x} = (x_i)$ be I dimensional frequency vector and $A = 1_I'$. In this case, $\boldsymbol{t}$ is the sample size n. This situation corresponds to testing the homogeneity of mean parameters for I independent Poisson variables conditional on the total sample size n. See also the example of Poisson regression below. In this case, a minimal Markov basis is formed as a set of $I-1$ degree 1 moves, but is not unique. A minimal Markov basis is constructed as follows. First consider the case of $n = |\boldsymbol{t}| = 1$. There are I elements in $\mathcal{F}_{\boldsymbol{t}}$ as

$$\mathcal{F}_{\boldsymbol{t}} = \{(1,0,\ldots,0)', (0,1,0,\ldots,0)', \ldots, (0,\ldots,0,1)'\}.$$

Each element $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{t}}$ forms an equivalence class by itself. To connect these points into a tree, there are I^{I-2} ways of choosing $I-1$ degree 1 moves by Cayley's theorem (see e.g. Chapter 4 of Wilson (1985)). One example is

$$\mathcal{B} = \{(1,-1,0,\ldots,0)', (0,1,-1,0,\ldots,0)', \ldots, (0,\ldots,0,1,-1)'\}.$$

It is easily verified that no move of degree larger than 1 is needed.

Two-way contingency tables. Next example is a standard two-way contingency table with fixed row and column sums. As is already seen, $\boldsymbol{x} = \{x_{ij}\}$ and

$$A = \begin{bmatrix} 1_I' \otimes E_J \\ E_I \otimes 1_J' \end{bmatrix}. \tag{8}$$

This is an elementary example of testing the hypothesis that the rows and the columns are independent. In this case, it is well known that the set of degree 2 moves displayed as

+1	−1
−1	+1

is a Markov basis. In addition, this is the unique minimal Markov basis from the discussion in the previous section. Indeed, for every $\boldsymbol{t}$ with $|\boldsymbol{t}| = 2$, except for a trivial case of one-element set $\#\mathcal{F}_{\boldsymbol{t}} = 1$, there are only two elements in $\mathcal{F}_{\boldsymbol{t}}$ and the above move is the difference of these two elements.

Three-way contingency tables with fixed two-dimensional marginals. Next we consider three-way contingency tables with fixed two-dimensional marginals. As we have seen in Section 2, $\boldsymbol{x} = \{x_{ijk}\}$ is the frequency vector of $I \times J \times K$ contingency table with lexicographical ordering of cells and A is written as

$$A = \begin{bmatrix} 1'_I \otimes E_J \otimes E_K \\ E_I \otimes 1'_J \otimes E_K \\ E_I \otimes E_J \otimes 1'_K \end{bmatrix}.$$

This corresponds to testing no three-way interactions of the log-linear model. As is already stated, it is surprisingly difficult to construct a connected Markov chain. Although an algebraic algorithm to calculate a Markov basis is given by Diaconis and Sturmfels (1998), any explicit characterization of a Markov basis is not known at present, except for some special cases. For the case of $2 \times J \times K$ tables, an explicit form of a Markov basis is given in Diaconis and Sturmfels (1998). Their basis is a set of degree $4, 6, \dots, \min\{J, K\}$ moves, where a typical degree $2n$ move is the following $2 \times n \times n$ move displayed as

$$\begin{array}{|cccccc|}\hline +1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & +1 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & & +1 & -1 & 0 \\ 0 & 0 & \cdots & 0 & +1 & -1 \\ -1 & 0 & \cdots & 0 & 0 & +1 \\ \hline \end{array} \quad \begin{array}{|cccccc|}\hline -1 & +1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & +1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & & -1 & +1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & +1 \\ +1 & 0 & \cdots & 0 & 0 & -1 \\ \hline \end{array}.$$

All the other degree $2n$ moves are obtained from this by permutations of indices or axes.

For the case of $3 \times 3 \times K$ tables, Aoki and Takemura (2002a) proves that a Markov basis is given as a set of the following four types of moves (and permutation of their indices and axes).

$$\text{degree 4 move :} \quad \begin{array}{|ccccc|}\hline +1 & -1 & 0 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline -1 & +1 & 0 & 0 & 0 \\ +1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

$$\text{degree 6 move :} \quad \begin{array}{|ccccc|}\hline +1 & -1 & 0 & 0 & 0 \\ -1 & 0 & +1 & 0 & 0 \\ 0 & +1 & -1 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline -1 & +1 & 0 & 0 & 0 \\ +1 & 0 & -1 & 0 & 0 \\ 0 & -1 & +1 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

$$\text{degree 8 move :} \quad \begin{array}{|ccccc|}\hline +1 & -1 & 0 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline -1 & 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & +1 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline 0 & +1 & -1 & 0 & 0 \\ 0 & -1 & 0 & +1 & 0 \\ 0 & 0 & +1 & -1 & 0 \\ \hline \end{array}$$

$$\text{degree 10 move :} \quad \begin{array}{|ccccc|}\hline +1 & -1 & 0 & 0 & 0 \\ -1 & +1 & 0 & -1 & +1 \\ 0 & 0 & 0 & +1 & -1 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline -1 & 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & +1 \\ \hline \end{array} \quad \begin{array}{|ccccc|}\hline 0 & +1 & -1 & 0 & 0 \\ 0 & -1 & 0 & +1 & 0 \\ 0 & 0 & +1 & -1 & 0 \\ \hline \end{array}$$

It is observed that $\mathcal{F}_{\boldsymbol{t}}$ is a two element set for each $\boldsymbol{t} = A\boldsymbol{z}$ of the above moves $\boldsymbol{z}$ for the $2 \times J \times K$ case and for the $3 \times 3 \times K$ case. Hence these moves constitute the unique minimal basis for respective cases.

Three-way contingency tables with fixed one-dimensional marginals. We now consider general three-way tables with fixed one-dimensional marginals. This corresponds to testing the independence model for three-way tables. With lexicographic ordering of indices, the matrix A is written as

$$A = \begin{bmatrix} 1_I' \otimes 1_J' \otimes E_K \\ 1_I' \otimes E_J \otimes 1_K' \\ E_I \otimes 1_J' \otimes 1_K' \end{bmatrix}.$$

In this case, we construct a minimal Markov basis as follows.

There are two obvious patterns of moves of degree 2. An example of moves of type I is

$$z_{111} = z_{222} = 1, \; z_{211} = z_{122} = -1,$$

with other elements being 0. For the case of $2 \times 2 \times 2$ table, this move can be displayed as follows

$$\boxed{\begin{array}{cc} +1 & 0 \\ 0 & -1 \end{array}} \quad \boxed{\begin{array}{cc} -1 & 0 \\ 0 & +1 \end{array}}\ .$$

All the other moves of type I are obtained by permutation of indices or the axes.

An example of moves of type II is

$$z_{111} = z_{122} = 1, \; z_{112} = z_{121} = -1,$$

with other elements being 0. For the case of $2 \times 2 \times 2$ table, this move can be displayed as follows

$$\boxed{\begin{array}{cc} +1 & -1 \\ -1 & +1 \end{array}} \quad \boxed{\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}}\ .$$

All the other moves of type II are obtained by permutation of indices or the axes. Let $\mathcal{B}^*$ be the set of type I and type II degree 2 moves. Then we have the following proposition.

Proposition 1 *$\mathcal{B}^*$ is a Markov basis for three-way contingency tables with fixed one-dimensional marginals.*

Proof. In this problem it is obvious that no degree 1 move is applicable to any frequency vector. Furthermore it is easy to verify that every degree 2 move is either of type I or type II. It remains to verify that for $|\boldsymbol{t}| \geq 3$, $\mathcal{F}_{\boldsymbol{t}}$ itself constitutes one $\mathcal{B}^*$-equivalence class. We can now apply the same argument used for $3 \times 3 \times K$ tables with fixed two-dimensional marginals in Aoki and Takemura (2002a). Suppose that for some $\boldsymbol{t}$, $\mathcal{F}_{\boldsymbol{t}}$ consists of more than one $\mathcal{B}^*$-equivalence classes. Let $\mathcal{F}_1, \mathcal{F}_2$ denote two different $\mathcal{B}^*$-equivalence classes. Choose $\boldsymbol{x} \in \mathcal{F}_1, \boldsymbol{y} \in \mathcal{F}_2$ such that

$$|\boldsymbol{z}| = |\boldsymbol{x} - \boldsymbol{y}| = \sum_{i,j,k} |x_{ijk} - y_{ijk}|$$

is minimized. Because $\boldsymbol{x}$ and $\boldsymbol{y}$ are chosen from different $\mathcal{B}^*$-equivalence classes, this minimum has to be positive. In the following we let $z_{111} > 0$ without loss of generality.

Case 1: Suppose that there exists a negative cell $z_{i_0 11} < 0, i_0 \geq 2$. Then because $\sum_{j,k} z_{i_0 jk} = 0$, there exists $(j,k), j+k > 2$, with $z_{i_0 jk} > 0$. Then the four cells

$$(1,1,1), \; (i_0,1,1), \; (i_0,j,k), \; (1,j,k)$$

are in the positions of either type I move or type II move. In either case we can apply a type I move or a type II move to $\boldsymbol{x}$ or $\boldsymbol{y}$ and make $|\boldsymbol{z}| = |\boldsymbol{x} - \boldsymbol{y}|$ smaller, which is a contradiction. This argument shows that $\boldsymbol{z}$ can not contain both positive and negative elements in any one-dimensional slice.

Case 2: Now we consider the remaining case, where no one-dimensional slice of $\boldsymbol{z}$ contains both positive and negative elements. Since $\sum_{j,k} z_{1jk} = 0$, there exists (j_1, k_1), $j_1, k_1 \geq 2$, such that $z_{1j_1k_1} < 0$. Similarly there exists (i_1, k_2), $i_1, k_2 \geq 2$, such that $z_{i_1 1 k_2} < 0$. Then the four cells

$$(1, j_1, k_1), (1, 1, k_1), (i_1, 1, k_2), (i_1, j_1, k_2)$$

are in the positions of a type II move (if $k_1 = k_2$) or a type I move (if $k_1 \neq k_2$) and we can apply a degree 2 move. By doing this $|\boldsymbol{z}| = |\boldsymbol{x} - \boldsymbol{y}|$ may remain the same, but now z_{11k_1} becomes negative and this case reduces to Case 1. Therefore Case 2 itself is a contradiction. Q.E.D.

We show in the following that $\mathcal{B}^*$ is not a minimal Markov basis. Let $\boldsymbol{z}$ be a degree 2 move and let $\boldsymbol{t} = A\boldsymbol{z}^+$. If $\boldsymbol{z}$ is a type II move, it is easy to verify that $\mathcal{F}_{\boldsymbol{t}}$ is a two-element set $\{\boldsymbol{z}^+, \boldsymbol{z}^-\}$. Therefore degree 2 moves of type II belong to each Markov basis. On the other hand, if $\boldsymbol{z}$ is a type I move, $\mathcal{F}_{\boldsymbol{t}}$ is a four-element set. For the $2 \times 2 \times 2$ case, let $\boldsymbol{t} = (z_{1\cdot\cdot}, z_{2\cdot\cdot}, z_{\cdot 1 \cdot}, z_{\cdot 2 \cdot}, z_{\cdot\cdot 1}, z_{\cdot\cdot 2})' = (1,1,1,1,1,1)'$. Then it follows

$$\mathcal{F}_{(1,1,1,1,1,1)'} = \left\{ \boxed{\begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & 0 \\ 0 & 1 \end{matrix}}\ ,\ \boxed{\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix}}\ ,\ \boxed{\begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}}\ ,\ \boxed{\begin{matrix} 0 & 0 \\ 0 & 1 \end{matrix}}\ \boxed{\begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix}} \right\}.$$

To connect these elements to a tree, only three moves of type I are needed. In the $2 \times 2 \times 2$ case, there are $4^{4-2} = 16$ possibilities, such as

$$\boxed{\begin{matrix} +1 & -1 \\ 0 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & 0 \\ -1 & +1 \end{matrix}}\ ,\ \boxed{\begin{matrix} 0 & +1 \\ -1 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & -1 \\ +1 & 0 \end{matrix}}\ ,\ \boxed{\begin{matrix} 0 & 0 \\ +1 & -1 \end{matrix}}\ \boxed{\begin{matrix} -1 & +1 \\ 0 & 0 \end{matrix}}$$

or

$$\boxed{\begin{matrix} +1 & -1 \\ 0 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & 0 \\ -1 & +1 \end{matrix}}\ ,\ \boxed{\begin{matrix} +1 & 0 \\ -1 & 0 \end{matrix}}\ \boxed{\begin{matrix} 0 & -1 \\ 0 & +1 \end{matrix}}\ ,\ \boxed{\begin{matrix} +1 & 0 \\ 0 & -1 \end{matrix}}\ \boxed{\begin{matrix} -1 & 0 \\ 0 & +1 \end{matrix}}$$

and so on. From these considerations, a minimal Markov basis for $I \times J \times K$ tables consists of

$$3 \binom{I}{2} \binom{J}{2} \binom{K}{2}$$

degree 2 moves of type I and

$$I \binom{J}{2} \binom{K}{2} + J \binom{I}{2} \binom{K}{2} + K \binom{I}{2} \binom{J}{2}$$

degree 2 moves of type II.

Poisson regression. Here we consider a simple example of Poisson regression discussed in Diaconis, Eisenbud and Sturmfels (1998). Let $\boldsymbol{x} = (x_0, x_1, \ldots, x_4)'$ and

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}.$$

Diaconis, Eisenbud and Sturmfels (1998) states that the set of degree 2 moves,

$$\begin{aligned}\mathcal{B} &= \{(1,-1,-1,1,0)',(1,-1,0,-1,1)',(0,1,-1,-1,1)',\\ &\quad (1,-2,1,0,0)',(0,1,-2,1,0)',(0,0,1,-2,1)'\}\end{aligned}$$

enables a connected chain. Indeed, the above basis is a minimal Markov basis but is not unique. To see this, consider $\mathcal{F}_{\boldsymbol{t}}$ with $|\boldsymbol{t}| = 2$. There are 9 possible values of $\boldsymbol{t}$ as

$$\boldsymbol{t}' = (2,0),(2,1),\ldots,(2,8).$$

For the case of $\boldsymbol{t}' = (2,0),(2,1),(2,7),(2,8)$, there is only one element in $\mathcal{F}_{\boldsymbol{t}}$ and we need not any move. For the case of $\boldsymbol{t}' = (2,2),(2,3),(2,5),(2,6)$, there are two elements in $\mathcal{F}_{\boldsymbol{t}}$, but for the case of $\boldsymbol{t}' = (2,4)$ there are three elements in $\mathcal{F}_{\boldsymbol{t}}$ as

$$\mathcal{F}_{(2,4)'} = \{(1,0,0,0,1)',(0,1,0,1,0)',(0,0,2,0,0)'\}.$$

The elements of the above $\mathcal{B}$ corresponds to the difference of the two elements in $\mathcal{F}_{\boldsymbol{t}}$, $\boldsymbol{t}' = (2,2),(2,3),(2,5),(2,6)$, and $\{(1,-1,0,-1,1)',(0,1,-2,1,0)'\}$, which connects the three elements in $\mathcal{F}_{(2,4)'}$ into a tree. This is not the only pair of moves to connect the three elements in $\mathcal{F}_{(2,4)'}$ to form a tree. There are three possibilities, i.e.,

$$\begin{aligned}\mathcal{B}^* &= \{(1,-1,-1,1,0)',(1,-1,0,-1,1)',(0,1,-1,-1,1)',\\ &\quad (1,-2,1,0,0)',(1,0,-2,0,1)',(0,0,1,-2,1)'\}\end{aligned}$$

and

$$\begin{aligned}\mathcal{B}^{**} &= \{(1,-1,-1,1,0)',(1,0,2,0,1)',(0,1,-1,-1,1)',\\ &\quad (1,-2,1,0,0)',(0,1,-2,1,0)',(0,0,1,-2,1)'\}\end{aligned}$$

are also minimal Markov bases.

Hardy-Weinberg model. Consider the case of

$$\boldsymbol{x} = (x_{11}, x_{12}, \ldots, x_{1I}, x_{22}, x_{23}, \ldots, x_{2I}, x_{33}, \ldots, x_{II})'$$

and $\boldsymbol{t} = (t_1, \ldots, t_I)'$ defined as

$$t_i = 2x_{ii} + \sum_{j \neq i} x_{ij}, \qquad i = 1, \ldots, I,$$

where $x_{ij} = x_{ji}$ for $i > j$. In this case, A is written as

$$A = (A_I \; A_{I-1} \; \cdots \; A_1), \quad A_k = \left(O_{k \times (I-k)} \; B_k'\right)',$$

where B_k is the following $k \times k$ square matrix

$$B_k = \begin{bmatrix} 2 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

This corresponds to the conditional test of the Hardy-Weinberg proportion. For this problem, Guo and Thompson (1992) construct a connected Markov chain. Their basis consists of three types of degree 2 moves, namely, type 0, type 1 and type 2. Here the term type refers to the number of nonzero diagonal cells in the move. The examples of the moves are displayed as

$$\text{type 0: } \begin{array}{|cccc|} \hline 0 & +1 & -1 & 0 \\ & 0 & 0 & -1 \\ & & 0 & +1 \\ & & & 0 \\ \hline \end{array}, \quad \text{type 1: } \begin{array}{|cccc|} \hline +1 & -1 & -1 & 0 \\ & 0 & +1 & 0 \\ & & 0 & 0 \\ & & & 0 \\ \hline \end{array}, \quad \text{type 2: } \begin{array}{|cccc|} \hline +1 & 0 & -2 & 0 \\ & 0 & 0 & 0 \\ & & +1 & 0 \\ & & & 0 \\ \hline \end{array}.$$

We show in the following that their basis is not minimal, and a minimal basis is not unique. Consider $\mathcal{F}_{\boldsymbol{t}}$ with $|\boldsymbol{t}| = 2$ for the above three types of moves. If $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$ for moves $\boldsymbol{z}$ of type 1 or type 2, there are two elements in $\mathcal{F}_{\boldsymbol{t}}$ and the move of type 1 or type 2 is the difference of these two elements. But if $\boldsymbol{t} = A\boldsymbol{z}^+ = A\boldsymbol{z}^-$ for a move $\boldsymbol{z}$ of type 0, there are three elements in $\mathcal{F}_{\boldsymbol{t}}$. Then to connect these three elements to form a tree, we can chose two moves to construct a minimal Markov basis. (There are three ways of doing this.) For example, consider the case of $I = 4$ and $\boldsymbol{t} = (1,1,1,1)'$. $\mathcal{F}_{(1,1,1,1)'}$ is written as

$$\mathcal{F}_{(1,1,1,1)'} = \left\{ \begin{array}{|cccc|} \hline 0 & 1 & 0 & 0 \\ & 0 & 0 & 0 \\ & & 0 & 1 \\ & & & 0 \\ \hline \end{array}, \begin{array}{|cccc|} \hline 0 & 0 & 1 & 0 \\ & 0 & 0 & 1 \\ & & 0 & 0 \\ & & & 0 \\ \hline \end{array}, \begin{array}{|cccc|} \hline 0 & 0 & 0 & 1 \\ & 0 & 1 & 0 \\ & & 0 & 0 \\ & & & 0 \\ \hline \end{array} \right\}.$$

To connect these three elements to a tree, any two of the following type 0 moves of degree 2,

$$\begin{array}{|cccc|} \hline 0 & +1 & -1 & 0 \\ & 0 & 0 & -1 \\ & & 0 & +1 \\ & & & 0 \\ \hline \end{array}, \begin{array}{|cccc|} \hline 0 & +1 & 0 & -1 \\ & 0 & -1 & 0 \\ & & 0 & +1 \\ & & & 0 \\ \hline \end{array}, \begin{array}{|cccc|} \hline 0 & 0 & -1 & +1 \\ & 0 & +1 & -1 \\ & & 0 & 0 \\ & & & 0 \\ \hline \end{array}$$

can be included in a minimal Markov basis. Accordingly, $I(I-1)(I-2)(I-3)/12$ moves of type 0, $I(I-1)(I-2)/2$ moves of type 1 and $I(I-1)/2$ moves of type 2 constitute a minimal Markov basis. The basis by Guo and Thompson (1992) is not minimal in the sense that all of $I(I-1)(I-2)(I-3)/8$ moves of type 0 are used in the algorithm proposed by them.

3.2 Some discussion

In the examples above we saw that for some problems minimal Markov basis is unique and for other problems it is not unique. Clearly this depends only on the properties of matrix A. But it seems very difficult to give a simple necessary and sufficient condition on A such that minimal Markov basis is unique. In integer programming literature (e.g. Schrijver (1986)), an important condition is the total unimodularity of the matrix A. We have seen that in the case of two-way contingency tables minimal Markov basis is unique and it is well known that A in (8) is totally unimodular. However in the simplest case of one-way tables minimal Markov basis is not unique and yet $A = (1, \ldots, 1)$ is obviously totally unimodular. This shows that total unimodularity is not directly related to uniqueness of minimal Markov basis. We should also mention that A for three-way tables with fixed two-dimensional marginals in (1) is not totally unimodular in general. In fact we have found a submatrix of A in (1) with determinant 2 by simple computer search.

As mentioned in Section 2.2, Theorem 1 is conceptually constructive, building up a minimal Markov basis from below. However it is computationally difficult to characterize the $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$ for large $|\boldsymbol{t}|$ as discussed in Diaconis, Eisenbud and Sturmfels (1998). If we could easily select representative elements from $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes $\mathcal{F}_{\boldsymbol{t}}$ for each $\boldsymbol{t}$, then a minimal Markov basis could be constructed as described in Theorem 1. Another question is to find a theoretical upper bound for n_0 such that $\mathcal{F}_{\boldsymbol{t}}$ itself constitutes one $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence class for all $\boldsymbol{t}$ with $|\boldsymbol{t}| \geq n_0$. By the Hilbert basis theorem existence of such an n_0 is guaranteed, but if we do not know some upper bound for n_0 we can not actually stop forming $\mathcal{B}_{|\boldsymbol{t}|-1}$-equivalence classes of $\mathcal{F}_{\boldsymbol{t}}$.

As mentioned at the end of Section 2.1 it is a subtle question to determine which moves of a minimal Markov basis $\mathcal{B}$ are needed for connecting $\mathcal{F}_{\boldsymbol{t}}$ for a given $\boldsymbol{t}$. Obviously we only need those elements of $\mathcal{B}$, that are applicable to at least one frequency vector of $\mathcal{F}_{\boldsymbol{t}}$. However the set of these move may not be minimal for connecting $\mathcal{F}_{\boldsymbol{t}}$ for a given $\boldsymbol{t}$. See the discussion on corner minors for two-way tables in Section 3 of Diaconis, Eisenbud and Sturmfels (1998). We study this question on two-way tables with structural zeros in our subsequent work (Aoki and Takemura (2002b)).

References

[1] Aoki, S. (2001). Network algorithm for the exact test of Hardy-Weinberg proportion for multiple alleles. Technical Report METR 01-06, Department of Mathematical Engineering and Information Physics, The University of Tokyo. Submitted for publication.

[2] Aoki, S. (2002). Improving path trimming in a network algorithm for Fisher's exact test in two-way contingency tables. To appear in *Journal of Statistical Computation and Simulation.*

[3] Aoki, S. and Takemura, A. (2002a). Minimal basis for connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals. Technical Report METR 02-02, Department of Mathematical Engineering and Information Physics, The University of Tokyo. Submitted for publication.

[4] Aoki, S. and Takemura, A. (2002b). Unique minimal basis for connected Markov chain for testing quasi-independence and quasi-symmetry in two-way contingency tables with structural zeros. In preparation.

[5] Bishop, Y. M. M., Fienberg, S. E. and Holland P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* The MIT Press, Cambridge.

[6] Diaconis, P., Eisenbud, D. and Sturmfels, B. (1998). Lattice walks and primary decomposition. in *Mathematical Essays in Honor of Gian-Carlo Rota.* Sagan, B. E. and Stanley, R. P., editors, pp. 173–193, Birkhauser, Boston.

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

[8] Dinwoodie, J. H. (1998). The Diaconis-Sturmfels algorithm and rules of succession. *Bernoulli*, **4**, pp. 401–410.

[9] Guo, S. W. and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrika*, **48**, pp. 361-372.

[10] Mehta, C. R. and Patel, N. R. (1983). A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *Journal of the American Statistical Association*, **78**, pp. 427–434.

[11] Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press, Oxford.

[12] Sakata, T. and Sawae, R. (2000). Echelon form and conditional test for three way contingency tables. in *Proceedings of the 10-th Japan and Korea Joint Conference of Statistics*, pp. 333–338.

[13] Schrijver, A. (1986). *Theory of Linear and Integer Programming.* Wiley, Chichester.

[14] Wilson, R. J. (1985). *Introduction to Graph Theory.* 3rd ed., Longman, New York.