Malaysian Journal of Mathematical Sciences 11(S) February: 53–72 (2017)
Special Issue: Conference on Agriculture Statistics 2015 (CAS 2015)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Algebraic Method for Independence Model of Two-Way Contingency Tables

Mohammed, N.F. *[1,3], Rakhimov, I.S. [1,2], and Shitan, M.[1,2]

[1] *Department of Mathematics,Faculty of Science, Universiti Putra Malaysia, Malaysia.*
[2] *Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia.*
[3] *Department of Mathematics, Faculty of Ibn AL Haitham Education, University of Baghdad, Iraq*

E-mail: nadia79math@yahoo.com
**Corresponding author*

ABSTRACT

The main purpose of the study is to propose an algebraic method to obtain the set of all independence models of $I \times J$ two-way contingency tables with the same row sums and column sums which is called fiber in algebraic statistics. This method involves solving a system of linear algebraic equations that only rely on row sums and column sums of the $I \times J$ two-way contingency table. The MATLAB software was used to solve this system. The effectiveness of the purposed method is illustrated by applying to a contingency table of agriculture teachers$'$ perception of secondary school agriculture.

Keywords: Contingency tables, Markov basis, Toric ideals, MATLAB.

1. Introduction

Algebraic statistics is a recent and a rapidly developing field and this term was generated by Pistone et al. (2000). It has two origins. The first one is in the work by Pistone and Wynn (1996) and the second one is in the work by Diaconis et al. (1998). The techniques of computational commutative algebra have applications in many areas like optimization see Sturmfels (1996), computational biology see Hosten et al. (2005), Pachter and Sturmfels (2004, 2005), and, of course, statistics see Pistone et al. (2000). Both algebraists and statisticians actively developed computational algebraic since the publication of Diaconis et al. (1998). Diaconis et al. (1998) defined the notion of Markov basis and proved the fundamental fact that a Markov basis corresponds to a set of binomial generators of a toric ideal.

Recently investigators have studied the effects of algebraic geometry representations of contingency tables. A fertile ground for the growth of algebraic statistics is provided by contingency tables and it is one of the most widely used by the group of researchers like Dobra and Fienberg (2001, 2003), Slavkovic and Fienberg (2004) and Dobra et al. (2009).

The main purpose of this paper is to propose a new algebraic method to find all $I \times J$ two-way contingency tables that have the same row sums and column sums and this is called fiber in algebraic statistics. To overcome this problem of table counting, a system of linear algebraic equations has been solved by MATLAB software which is based only on the same row sums and column sums of the contingency table. The advantage of this method is that it only depends on the dimension, the row sums and column sums of the contingency table. Furthermore, it can be implemented easily and effectively.

The paper is organized as follows. In Section 2, some notations on contingency tables, sufficient statistics, fiber, Markov basis and toric ideal are given. In Section 3, we show how to compute a fiber by using our algebraic method. Then, we apply the method to a numerical example in Section 4. We end the paper with conclusion in Section 5.

2. Preliminaries

In this section, we review some relevant concepts needed for the work.

Let Γ be a finite set with $p =\mid \Gamma \mid$ elements, we call an element of Γ a cell and it$^{'}$s represented by $\mathbf{i} \in \Gamma$. In the case of $I \times J$ two-way contingency table

where I is a row factor indexed by $\mathbf{i}$ and J is a column factor indexed by j. The cell $\mathbf{i}$ is often a multi-index $\mathbf{i} = (i, j)$ and $p = I \times J$. A non-negative integer $x_{\mathbf{i}} \in \mathbb{N} = \{0, 1, \cdots\}$ represent the frequency of a cell $\mathbf{i}$. The set of frequencies is known as **contingency table** and it is represented by $\mathbf{x} = \{x_{\mathbf{i}}\}_{\mathbf{i} \in \Gamma}$.
A contingency table $\mathbf{x} = \{x_i\}_{i \in \Gamma}$ can be treated as a p -dimensional column vector of non-negative integers with an appropriate ordering of the cells. Additionally, it can also be considered as a function from Γ to $\mathbb{N}$ define as $\mathbf{i} \longmapsto x_{\mathbf{i}}$. The L_1-norm of $\mathbf{x} \in \mathbb{N}^p$ is known as the sample size and is represented by $n = \sum_{\mathbf{i} \in \Gamma} x_{\mathbf{i}}$. Let $\mathbb{Z}$ denote the set of integer numbers and $\mathbf{a}_j \in \mathbb{Z}^p, j = 1, \cdots, v$, where $v = I + J$, as fixed column vectors comprising of integers. A v- dimensional column vector $\mathbf{t} = (t_1, \cdots, t_v)^{'} \in \mathbb{Z}^v$ is defined as $t_j = \mathbf{a}_j^{'} x, j = 1, \cdots, v$. In this context, $'$ represents the transpose of a vector or matrix. We also define a $v \times p$ matrix A, with it's j-row being $\mathbf{a}_j^{'}$ given by $A = \begin{bmatrix} a_1^{'} \\ \vdots \\ a_v^{'} \end{bmatrix}$ and $\mathbf{t} = A\mathbf{x}$ is a v- dimensional column vector.

The kernel of a linear map $A : \mathbb{Q}^p \longrightarrow \mathbb{Q}^v$ between two vector spaces $\mathbb{Q}^p$ and $\mathbb{Q}^v$ over rational numbers $\mathbb{Q}$ is defined as $kerA = \{\mathbf{y} \in \mathbb{Q}^p : A\mathbf{y} = \mathbf{0}\}$, where $\mathbf{0}$ denotes the zero vector in $\mathbb{Q}^v$. The rank of A is given as $rankA = I + J - 1$ and the dimension of the kernel of A is given as dim $kerA = IJ - rankA$. In typical situations of a statistical theory, $\mathbf{t}$ is sufficient statistic for the nuisance parameter and the set of $x^{'}$s for a given $\mathbf{t}$, $F_{\mathbf{t}} = \{\mathbf{x} \in \mathbb{N}^p : A\mathbf{x} = \mathbf{t}\}$, is considered to perform similar tests. In the case of two$-$way contingency table which is an independent model, $\mathbf{t}$ is a v- dimensional column vector of the row sums and column sums of $\mathbf{x}$, and $F_{\mathbf{t}}$ is the set of $\mathbf{x}^{'}$ s with the same row sums and column sums to $\mathbf{t}$, we call $F_{\mathbf{t}}$ a **t-fiber**. In fact, if we define $\mathbf{x}_1 \sim \mathbf{x}_2 \Longleftrightarrow \mathbf{x}_1 - \mathbf{x}_2 \in ker(A)$, then $\sim$ is an equivalence relation and $\mathbb{N}^p$ is partitioned into disjoint equivalence classes.Furthermore, $\mathbf{t}$ may be considered as labels of these equivalence classes,see Aoki and Takemura (2008), Drton et al. (2008). A p-dimensional column vector of integers $\mathbf{z} = \{z_{\mathbf{i}}\}_{\mathbf{i} \in \Gamma} \in \mathbb{Z}^p$ is known as a **move** if it is in the kernel of A, i.e. $A\mathbf{z} = \mathbf{0}$. The integer kernel of A is the set of moves (for a given A) and denote by $ker_{\mathbb{Z}}A = \mathbb{Z}^p \cap ker(A)$.

Let $B \subset ker_{\mathbb{Z}}A$ be a finite set of moves for a configuration A.The set B is called a **Markov basis** if for all fiber $F_{\mathbf{t}}$ and for all elements $\mathbf{x}_1, \mathbf{x}_2 \in F_{\mathbf{t}}, x_1 \neq x_2$, there exist $K > 0$,$\mathbf{z}_1, \cdots, \mathbf{z}_k \in B$ and $\epsilon_1, \cdots, \epsilon_k \in \{-1, +1\}$, such that

$$\mathbf{x_2} = \mathbf{x_1} + \sum_{k=1}^{K} \epsilon_k \mathbf{z}_k, \qquad \mathbf{x_1} + \sum_{k=1}^{L} \epsilon_k \mathbf{z}_k \in F_{\mathbf{t}}, \qquad L = 1, \cdots, K-1 \tag{1}$$

Then we call x_2 is accessible from x_1 by B and this is denoted by $\mathbf{x}_1 \sim \mathbf{x}_2(\text{mod } B)$. Therefore, by adding or subtracting moves from B we can move all over any fiber if Markov basis B is given.
Now consider the following integer matrix $\mathbf{z} = \mathbf{z}(i_1, i_2; j_1, j_2) = \{z_{ij}\}$, defined by

$$z_{ij} = \begin{cases} +1 & (i,j) = (i_1, j_1), (i_2, j_2), \\ -1 & (i,j) = (i_1, j_2), (i_2, j_1), \\ 0 & otherwise \end{cases} \tag{2}$$

where i_1 and i_2 are any two rows and j_1 and j_2 are any columns from $I \times J$ two-way contingency table, respectively.

Let $B = \{z(i_1, i_2; j_1, j_2) : 1 \leqslant i_1 < i_2 \leqslant I, 1 \leqslant j_1 < j_2 \leqslant J\}$. Then B forms a Markov basis for the $I \times J$ independence model of two-way contingency tables.

For contingency tables, the above notations can be translated to the objects on polynomial rings. Let $\mathbf{u} = \{u_i\}_{i \in \Gamma}$ be the set of indeterminates and let $K[\mathbf{u}]$ represent the polynomial ring in the indeterminate $\mathbf{u}$ over a field K. We often denote u_i as $u(\mathbf{i})$. Then a contingency table $\mathbf{x} = \{x_i\}_{i \in \Gamma}$ is specified as a monomial

$$\mathbf{u}^x = \prod_{i \in \Gamma} u(\mathbf{i})^{x_{\mathbf{i}}} \in k[\mathbf{u}] \tag{3}$$

and as a binomial

$$f(\mathbf{u}) = \mathbf{u}^{\mathbf{y}} - \mathbf{u}^{\mathbf{x}} = \prod_{i \in \Gamma} u(\mathbf{i})^{y_{\mathbf{i}}} - \prod_{i \in \Gamma} u(\mathbf{i})^{x_{\mathbf{i}}} \tag{4}$$

such that $\mathbf{x}, \mathbf{y} \in F_t$ and $\mathbf{y} - \mathbf{x} \in ker_{\mathbb{Z}}A$.

One of the main object considered in algebraic statistics is toric ideal. To define the toric ideal, let us introduce indeterminates $q_1, ..., q_v$ corresponding to the rows of A and $u_1, \cdots, u_p$ corresponding to the columns of matrix A such that $\mathbf{q} = \{q_1, \cdots, q_v\}$ and $\mathbf{u} = \{u_1, \cdots, u_p\}$. Let $K[\mathbf{u}]$ and $K[\mathbf{q}]$ denote the polynomial rings in the indeterminates $\mathbf{u}$ and $\mathbf{q}$ over a field K, respectively. Considering a map π_A from $K[\mathbf{u}]$ to $K[\mathbf{q}]$ such that each indeterminate $u(\mathbf{i})$ is mapped to a monomial in $K[\mathbf{q}]$ as

$$\pi_A : u(\mathbf{i}) \mapsto \mathbf{q}^{a(\mathbf{i})} = q_1^{a_1(\mathbf{i})} q_2^{a_2(\mathbf{i})} \cdots q_v^{a_v(\mathbf{i})} \tag{5}$$

where $a(\mathbf{i})$ is the ith column of A.

The **toric ideal** $I_A =< \{f \in K[\mathbf{u}] : \pi_A(f) = \mathbf{0}\} >$, for a polynomial $f \in K[\mathbf{u}]$, is the kernel of π_A and it is the ideal generated by binomials $f = (\mathbf{u}^{\mathbf{y}} - \mathbf{u}^{\mathbf{x}})$ such that $\mathbf{y} - \mathbf{x} \in ker_Z A$.The relationship between moves and binomials of a toric ideal is described in the fundamental theorem of Markov basis which state that a finite set of moves B is a Markov basis for A if and only if the set of binomials $\{\mathbf{u}^{\mathbf{y}} - \mathbf{u}^{\mathbf{x}} : \mathbf{y} - \mathbf{x} = \mathbf{z} \in B\}$ generates the toric ideal I_A , see Aoki et al. (2012).

Another main object considered in algebraic statistics is a graph. A graph is a pair $G = (V, E)$ of sets, where the vertex set of a graph G is referred to as $V(G)$, its edges set as $E(G)$. We call a graph G is connected if for every pair of distinct vertices $u, v \in V(G)$ has a path from u to v. Otherwise, we say the graph is disconnected Agnarsson and Greenlaw (2006). To construct a connected graph, let $\mathbf{x}_2$ be accessible from $\mathbf{x}_1$ by B i.e. $\mathbf{x}_1 \sim \mathbf{x}_2 \pmod{B}$. Obviously, the accessibility by B is an equivalence relation and each fiber $F_{\mathbf{t}}$ is partitioned into disjoint equivalence classes by moves of B. Furthermore, the equivalence classes are called B -equivalence classes of fiber $F_{\mathbf{t}}$. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be elements from two different B -equivalence classes of fiber $F_{\mathbf{t}}$. Then, a move $\mathbf{z} = \mathbf{x}_1 - \mathbf{x}_2$ connects these two equivalence classes. All B -equivalence classes can be connected by creating an undirected graph $G_{(\mathbf{t},B)}$, where B -equivalence classes of fiber $F_{\mathbf{t}}$ are interpreted as vertices of the graph $G_{(\mathbf{t},B)}$ and connecting moves are interpreted as edges of an undirected graph $G_{(\mathbf{t},B)}$. If $G_{(\mathbf{t},B)}$ is a connected graph for all $\mathbf{t}$ with $F_t \neq \emptyset$, then B is Markov basis. Therefore, to be able to move all over $F_{\mathbf{t}}$, it should correspond to the connectedness of $G_{(\mathbf{t},B)}$, see Takemura and Aoki (2004).

Finally, we need to illustrate some terminology that will be needed for solving systems of equations. A matrix is said to be in row reduced echelon form (RREF) if all nonzero rows are above any zero row, the first nonzero entry in a row (the leading entry) is a one and every other entry in a column with a leading one is zero. Those columns with a leading entry are known as pivot columns, and the leading entries are called pivot positions. A free variable of a linear system is a variable which is associated with a column in the RREF matrix which is not a pivot column.

3. The Proposed Method

In this section, we explain our proposed method based on row sums and column sums of independence model of $I \times J$ two-way contingency tables for solving a nonhomogeneous system of linear algebraic equations. Firstly, we roughly characterize the theoretical basis of the method. Supposing, we intend to solve the following non-homogeneous linear system of equations:

$$A\mathbf{x} = \mathbf{t} \tag{6}$$

where A is a configuration matrix and has size $v \times p$, x is a contingency table and can be written as p - dimensional column frequency vector and $\mathbf{t}$ is a v-dimensional column vector.

In the case of the independence model of two−way contingency tables, $\mathbf{t}$ will be a v- dimensional column vector of the row sums and column sums of a contingency table $\mathbf{x}$.

This system consists of a set of $(I+J-1)$ linear equations in $(I \times J)$ variables. It is clear that there are fewer equations than variables. Therefore, this system is regarded underdetermined system. Let $x_{ij}, i = 1, 2, \cdots, I, j = 1, 2, \cdots, J$, be the observed cell. Furthermore, let the leading variables be the frequency of cells in the first row and in the first column denoted as x_{11}, x_{k1}, x_{1l} and the free variables be the other frequency of cells in the contingency table $\mathbf{x}$ and denoted by

$$x_{kl} = d_{(k-1)(l-1)}, \qquad k = 2, \cdots, I, \quad l = 2, \cdots, J \tag{7}$$

.

The number of leading variables is $I+J-1$ and the number of free variables is $IJ-(I+J-1)$. However, $rankA = I+J-1$ and dim $kerA = IJ - rankA$. Therefore, the number of leading variables equals $rankA$ and the number of free variables equals dim $kerA$.

The row sums and the column sums are presented as

$$r_i = \sum_{j=1}^{J} x_{ij}, \quad i = 1, 2, \cdots, I, \tag{8}$$

and

$$c_j = \sum_{i=1}^{I} x_{ij}, \qquad j = 1, 2, \cdots, J \tag{9}$$

respectively. It is clear that $I \times J$ two-way contingency table $\mathbf{x}$ has I rows and J columns, see Table 1. Let r be a I- dimension column vector of all row sums of a contingency table $\mathbf{x}$ and represented as $\mathbf{r} = (r_1, \cdots, r_I)^{'}$. Let c represent a J-dimension column vector of all column sums of a contingency table $\mathbf{x}$ and can be represented as $\mathbf{c} = (c_1, \cdots, c_J)^{'}$. We also define a $(I-1) \times (J-1)$ matrix d of all free variables of a contingency table. We treat a matrix d as a column vector and represented as $d = (d_{11}, d_{12}, \cdots, d_{1(J-1)}, d_{21}, \cdots, d_{2(J-1)}, \cdots, d_{(I-1)1}, \cdots, d_{(I-1)(J-1)})^{'}$. Then, we find the infimum and supremum (abbreviated inf (sup)) for each free variables $x_{kl} = d_{(k-1)(l-1)}$ in the contingency table $\mathbf{x}$, such that

$$inf(d_{(k-1)(l-1)}) = 0 \tag{10}$$

and

$$sup(d_{(k-1)(l-1)}) = min(r_k, c_l) \tag{11}$$

where

$$r_k = x_{k1} + \sum_{j=2}^{J} x_{kl}, \quad c_l = x_{1l} + \sum_{i=2}^{I} x_{kl}, \qquad k = 2, \cdots, I, \quad l = 2, \cdots, J \tag{12}$$

respectively.

Furthermore, we find the constraints for the free variables where

$$\sum_{l=2}^{J} d_{(k-1)(l-1)} \leq r_k, \quad \sum_{k=2}^{I} d_{(k-1)(l-1)} \leq c_l, \tag{13}$$

$$\sum_{k=2}^{I}\sum_{l=2}^{J} d_{(k-1)(l-1)} \leq min(\sum_{k=2}^{I} r_k, \sum_{l=2}^{J} c_l), \tag{14}$$

and

$$\sum_{k=2}^{I}\sum_{l=2}^{J} d_{(k-1)(l-1)} \geq \sum_{k=2}^{I} r_k - \sum_{i=1}^{I} x_{il}. \tag{15}$$

Finally, we find the equations of dependent variables and free variables that represented as

$$x_{11} = c_1 - \sum_{k=2}^{I} r_k + \sum_{k=2}^{I}\sum_{l=2}^{J} d_{(k-1)(l-1)}, \tag{16}$$

$$x_{k1} = r_k - \sum_{l=2}^{J} d_{(k-1)(l-1)}, \tag{17}$$

$$x_{1l} = c_l - \sum_{k=2}^{I} d_{(k-1)(l-1)}, \tag{18}$$

$$x_{kl} = d_{(k-1)(l-1)} \tag{19}$$

.

Table 1: $I \times J$ Two-Way Contingency Table to Show the Leading Variables and Free Variables

x_{11}	x_{12}	$\cdots$	x_{1J}	r_1
x_{21}	x_{22}=d_{11}	$\cdots$	x_{2J}=$d_{1(J-1)}$	r_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{I1}	x_{I2} =$d_{(I-1)2}$	$\cdots$	$x_{IJ} = d_{(I-1)(J-1)}$	r_I
c_1	c_2	$\cdots$	c_J	n

Thus, the proposed method is summarized as follows:

Step 1: Input the number of rows I, the number of column J, the vector of all row sums r, the vector of all column sums c.

Step 2: For $d_{11} = 0$ to $\min\,(r_2, c_2)$; $d_{12} = 0$ to $\min\,(r_2, c_3)$; $\ldots$; $d_{(I-1)(J-1)} = 0$ to $\min(r_I, c_J)$.

Step 3: If $\sum\limits_{l=2}^{J} d_{(k-1)(l-1)} \leq r_k, \sum\limits_{k=2}^{I} d_{(k-1)(l-1)} \leq c_l, \sum\limits_{k=2}^{I}\sum\limits_{l=2}^{J} d_{(k-1)(l-1)} \leq \min(\sum\limits_{k=2}^{I} r_k, \sum\limits_{l=2}^{J} c_l)$ and $\sum\limits_{k=2}^{I}\sum\limits_{l=2}^{J} d_{(k-1)(l-1)} \geq \sum\limits_{k=2}^{I} r_k - \sum\limits_{i=1}^{I} x_{il}$ then go to step 4. If $\sum\limits_{l=2}^{J} d_{(k-1)(l-1)} > r_k$ or $\sum\limits_{k=2}^{I} d_{(k-1)(l-1)} > c_l$ or if $\sum\limits_{k=2}^{I}\sum\limits_{l=2}^{J} d_{(k-1)(l-1)} \geq \min(\sum\limits_{k=2}^{I} r_k, \sum\limits_{l=2}^{J} c_l)$ or $\sum\limits_{k=2}^{I}\sum\limits_{l=2}^{J} d_{(k-1)(l-1)} \leq \sum\limits_{k=2}^{I} r_k - \sum\limits_{i=1}^{I} x_{il}$ then go to step 2.

Step 4: $x_{11} = c_1 - (\sum\limits_{k=2}^{I} r_k) + \sum\limits_{k=2}^{I}\sum\limits_{l=2}^{J} d_{(k-1)(l-1)}$,
$x_{k1} = r_k - \sum\limits_{l=2}^{J} d_{(k-1)(l-1)}$, $x_{1l} = c_l - \sum\limits_{k=2}^{I} d_{(k-1)(l-1)}$, $x_{kl} = d_{(k-1)(l-1)}$.

4. Application

In this section, we apply our method to a 3 × 2 two-way contingency table which consists of agriculture teachers′ perception of secondary school agriculture to show the effectiveness of the purposed method.

By this example, it is illustrated that our method can be easily implemented in MATLAB. Table 2 shows agriculture data gathered to test the hypothesis of the association between agriculture′s professional qualifications and their perception. The table enrollment thirty- one agriculture teachers were chosen from a random sampling of agriculture schools to determine the relationship between agriculture teacher′s professional qualification and perception of secondary school agriculture. The agriculture respondents were asked to state their professional qualifications which were divided into three groups: technically and professionally, technically trained and others. To be able to measure perception, the teachers were asked to respond to items about secondary school agriculture. Every item in the questionnaire was rated on a scale of five points (Strongly Agree, Agree, Uncertain, Disagree and Strongly Disagree). This was used to calculate a mean rating score for all teachers. Each individual′s mean rating score on perception was classified as either high or low depending on whether it was an above or below the mean rating score for the group, see Muchiri and Kiriungi (2013).

Table 2: : 3 × 2 : Two-Way Contingency Table to Show the Relationship between Agriculture′s Professional Qualifications and their Perception of Secondary School Agriculture

	Perception	
Professional Qualifications	Low	High
Technically and professionally trained	7	4
Technically trained	6	7
Others	2	5
Total	15	16

To test the independence between the perception of secondary school agriculture and the agriculture teachers′ professional qualifications, let the teachers′ professional qualifications and their perception are regarded as discrete random variables X and Y with possible values X_i and Y_j, where i runs from $1, 2, 3$ and j runs from $1, 2$ respectively. The joint probability distribution of X and Y is represented as

$$p_{ij} = P(X = i, Y = j), \quad i = 1, 2, 3, j = 1, 2. \tag{20}$$

and let the odds ratio is defined to be

$$OR = \frac{(p_{ij}p_{kl})}{(p_{il}p_{kj})} \quad forall \quad 1 \leq i < k \leq I, \quad 1 \leq j < l \leq J. \tag{21}$$

The hypothesises of independence can be stated as

H_0: There is no significant relationship between agriculture teacher$^{'}$s professional qualifications and their perception of secondary school agriculture.

H_1: There is a relationship between agriculture teacher$^{'}$s professional qualifications and their perception of secondary school agriculture.
To evaluate the relationship between agriculture$^{'}$s professional qualifications and their perception of secondary school agriculture, odds ratio test was applied and found the odds ratio $OR = 1$. Therefore, it led to the acceptance of the null hypothesis because the two random variables X and Y are independent when the odds ratio $OR = 1$, see Agresti (1996), Agresti and Kateri (2011).

4.1 Two-Way Contingency Tables with the Same Row Sums and Column Sums

We applied our algebraic method on 3×2 two-way contingency table of agriculture teachers$^{'}$ perception of secondary school agriculture by using MATLAB language, see Table 3, to find the fiber.

Step1: $I = 3, J = 2, r = (11, 13, 7), c = (15, 16)$.

Step 2: For $d_{11} = 0\colon \min(13, 16), d_{21} = 0\colon \min(7, 16)$.

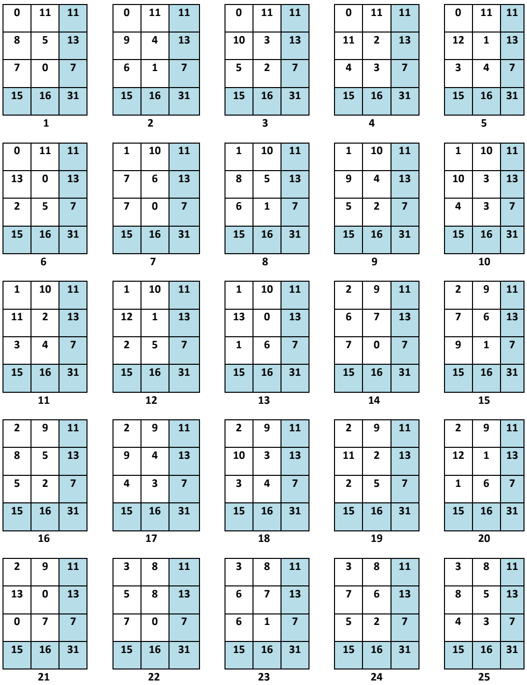
Step 3: If $d_{11} + d_{21} \leq \min\ (20, 16)$ and $d_{11} + d_{21} \geq 5$ then go to step 4. If $d_{11} + d_{21} \geq \min\ (20, 16)$ or $d_{11} + d_{21} \leq 5$ then go to step 2.

Step 4: $x_{11} = d_{11} + d_{21} - 5, x_{12} = 16 - (d_{11} + d_{21}), x_{21} = 13 - d_{11}, x_{31} = 7 - d_{21}, x_{22} = d_{11}, x_{32} = d_{21}$.

When we apply these steps of our method in MATLAB, the result will be 87 contingency tables with the same rows sums and columns sums see Figure $1, 2, 3$ and 4.

Table 3: :$I \times J$ Two-Way Contingency Table to Show the Leading Variables and Free Variables

x_{11}	x_{12}	11
x_{21}	$x_{22} = d_{11}$	13
x_{31}	$x_{32} = d_{21}$	7
15	16	31



0	11	11
8	5	13
7	0	7
15	16	31

1

0	11	11
9	4	13
6	1	7
15	16	31

2

0	11	11
10	3	13
5	2	7
15	16	31

3

0	11	11
11	2	13
4	3	7
15	16	31

4

0	11	11
12	1	13
3	4	7
15	16	31

5

0	11	11
13	0	13
2	5	7
15	16	31

6

1	10	11
7	6	13
7	0	7
15	16	31

7

1	10	11
8	5	13
6	1	7
15	16	31

8

1	10	11
9	4	13
5	2	7
15	16	31

9

1	10	11
10	3	13
4	3	7
15	16	31

10

1	10	11
11	2	13
3	4	7
15	16	31

11

1	10	11
12	1	13
2	5	7
15	16	31

12

1	10	11
13	0	13
1	6	7
15	16	31

13

2	9	11
6	7	13
7	0	7
15	16	31

14

2	9	11
7	6	13
9	1	7
15	16	31

15

2	9	11
8	5	13
5	2	7
15	16	31

16

2	9	11
9	4	13
4	3	7
15	16	31

17

2	9	11
10	3	13
3	4	7
15	16	31

18

2	9	11
11	2	13
2	5	7
15	16	31

19

2	9	11
12	1	13
1	6	7
15	16	31

20

2	9	11
13	0	13
0	7	7
15	16	31

21

3	8	11
5	8	13
7	0	7
15	16	31

22

3	8	11
6	7	13
6	1	7
15	16	31

23

3	8	11
7	6	13
5	2	7
15	16	31

24

3	8	11
8	5	13
4	3	7
15	16	31

25

Figure 1

3	8	11
9	4	13
3	4	7
15	16	31

26

3	8	11
10	3	13
2	5	7
15	16	31

27

3	8	11
11	2	13
1	6	7
15	16	31

28

3	8	11
12	1	13
0	7	7
15	16	31

29

4	7	11
4	9	31
7	0	7
15	16	31

30

4	7	11
5	8	13
6	1	7
15	16	31

31

4	7	11
6	7	13
5	2	7
15	16	31

32

4	7	11
7	6	13
4	3	7
15	16	31

33

4	7	11
8	5	13
3	4	7
15	16	31

34

4	7	11
9	4	13
2	5	7
15	16	31

35

4	7	11
10	3	13
1	6	7
15	16	31

36

4	7	11
11	2	13
0	7	7
15	16	31

37

5	6	11
3	10	13
7	0	7
15	16	31

38

5	6	11
4	9	13
6	1	7
15	16	31

39

5	6	11
5	8	13
5	2	7
15	16	31

40

5	6	11
6	7	13
4	3	7
15	16	31

41

5	6	11
7	6	13
3	4	7
15	16	31

42

5	6	11
8	5	13
2	5	7
15	16	31

43

5	6	11
9	4	13
1	6	7
15	16	31

44

5	6	11
10	3	13
0	7	7
15	16	31

45

6	5	11
2	11	13
7	0	7
15	16	31

46

6	5	11
3	10	13
6	1	7
15	16	31

47

6	5	11
4	9	13
5	2	7
15	16	31

48

6	5	11
5	8	13
4	3	7
15	16	31

49

6	5	11
6	7	13
3	2	7
15	16	31

50

Figure 2

6	5	11
7	6	13
2	5	7
15	16	31

51

6	5	11
8	5	13
1	6	7
15	16	31

52

6	5	11
9	4	13
0	7	7
15	16	31

53

7	4	11
1	12	13
7	0	7
15	16	31

54

7	4	11
2	11	13
6	1	7
15	16	31

55

7	4	11
3	10	13
5	2	7
15	16	31

56

7	4	11
4	9	13
4	3	7
15	16	31

57

7	4	11
5	8	13
3	4	7
15	16	31

58

7	4	11
6	7	13
2	5	7
15	16	31

59

7	4	11
7	6	13
1	6	7
15	16	31

60

7	4	11
8	5	13
0	7	7
15	16	31

61

8	3	11
0	13	13
7	0	7
15	16	31

62

8	3	11
1	12	13
6	1	7
15	16	31

63

8	3	11
2	11	13
5	2	7
15	16	31

64

8	3	11
3	10	13
4	3	7
15	16	31

65

8	3	11
4	9	13
3	4	7
15	16	31

66

8	3	11
5	8	13
2	5	7
15	16	31

67

8	3	11
6	7	13
1	6	7
15	16	31

68

8	3	11
7	6	13
0	7	7
15	16	31

69

9	2	11
0	13	13
6	1	7
15	16	31

70

9	2	11
1	12	13
5	2	7
15	16	31

71

9	2	11
2	11	13
4	3	7
15	16	31

72

9	2	11
3	10	13
3	4	7
15	16	31

73

9	2	11
4	9	13
2	5	7
15	16	31

74

9	2	11
5	8	13
1	6	7
15	16	31

75

Figure 3

9	2	11
6	7	13
0	7	7
15	16	31

76

10	1	11
0	13	13
5	2	7
15	16	31

77

10	1	11
1	12	13
4	3	7
15	16	31

78

10	1	11
2	11	13
3	4	7
15	16	31

79

10	1	11
3	10	13
2	5	7
15	16	31

80

10	1	11
4	9	13
1	6	7
15	16	31

81

10	1	11
5	8	13
0	7	7
15	16	31

82

11	0	11
0	13	13
4	3	7
15	16	31

83

11	0	11
1	12	13
3	4	7
15	16	31

84

11	0	11
2	11	13
2	5	7
15	16	31

85

11	0	11
3	10	13
1	6	7
15	16	31

86

11	0	11
4	9	13
0	7	7
15	16	31

87

Figure 4

4.2 Toric Ideal and Markov Bases

To find the toric ideal I_A for 3×2-contingency tables, we will define a map π_A from $k[\mathbf{u}]$ to $k[\mathbf{q}]$, where $\mathbf{u} = u_1, u_2, u_3, u_4, u_5, u_6$ and $\mathbf{q} = q_1, q_2, q_3, q_4, q_5$ such that each $u(\mathbf{i})$ is mapped to a monomial in $k[\mathbf{q}]$ as

$$\pi_A(u(\mathbf{i})) = \mathbf{q}^{a(i)} = q_1^{a_1(\mathbf{i})} q_2^{a_2(\mathbf{i})} \cdots q_5^{a_5(\mathbf{i})} \tag{22}$$

.
Then, for a monomial $\mathbf{u}^{\mathbf{x}}$,

$$\pi_A(\mathbf{u}^{\mathbf{x}}) = \pi_A(\prod_{\mathbf{i}\in\Gamma} u(\mathbf{i})^{x(\mathbf{i})}) = \prod_{\mathbf{i}\in\Gamma} \pi_A(u(\mathbf{i}))^{x(\mathbf{i})} = \prod_{\mathbf{i}\in\Gamma} \mathbf{q}^{a(\mathbf{i})x(\mathbf{i})} = \prod_{j=1}^{v=5} q_j^{\sum_{\mathbf{i}\in\Gamma} a(\mathbf{i})x(\mathbf{i})} = \mathbf{q}^{A\mathbf{x}} \tag{23}$$

For example, if we consider the contingency table $x_{59} = (7, 4, 6, 7, 2, 5)$. Then for a monomial $\mathbf{u}^{\mathbf{x}} = u_1^7 u_2^4 u_3^6 u_4^7 u_5^2 u_6^5$ and $\pi_A(\mathbf{u}^{\mathbf{x}}) = \pi_A(u_1^7)\pi_A(u_2^4) \cdots \pi_A(u_6^5) = (q_1q_4)^7(q_1q_5)^4 \cdots (q_3q_5)^5 = q_1^{11} q_2^{13} q_3^7 q_4^{15} q_5^{16} = \mathbf{q}^{A\mathbf{x}} = \mathbf{q}^{\mathbf{t}}$.

Let us illustrate π_A for the case of the independence model of two-way tables under the multinomial sampling scheme. Let (i, j) denote the cell of a two-way table and consider the probability p_{ij} of the cell as an indeterminate (instead of $u(\mathbf{i})$). Under the independence model $p_{ij} = r_i c_j$. We can understand this by substituting $r_i c_j$ into p_{ij} and consider $\pi_A : p_{ij} \to r_i c_j$. For $I = 3, J = 2$, let $\mathbf{u} = \{p_{11}, p_{12}, p_{21}, p_{22}, p_{31}, p_{32}\}$ and $\mathbf{q} = \{r_1, r_2, r_3, c_1, c_2\}$, for example, we consider the two contingency table $\mathbf{x} = \mathbf{x}_{59} = (7, 4, 6, 7, 2, 5)$ and $\mathbf{y} = \mathbf{x}_{60} = (7, 4, 7, 6, 1, 6)$. Then

$$\pi_A(f) = \pi_A(\mathbf{u}^{\mathbf{y}} - \mathbf{u}^{\mathbf{x}}) = \pi_A(\mathbf{u}^{\mathbf{y}}) - \pi_A(\mathbf{u}^{\mathbf{x}}) =$$

$$\pi_A(p_{11}^7 p_{12}^4 p_{21}^7 p_{22}^6 p_{31}^1 p_{32}^6) - \pi_A(p_{11}^7 p_{12}^4 p_{21}^6 p_{22}^7 p_{31}^2 p_{32}^5) =$$

$$\begin{gathered}(r_1c_1)^7(r_1c_2)^4(r_2c_1)^7(r_2c_2)^6(r_3c_1)^1(r_3c_2)^6 - \\ (r_1c_1)^7(r_1c_2)^4(r_2c_1)^6(r_2c_2)^7(r_3c_1)^2(r_3c_2)^5 =\end{gathered}$$

$$r_1^{11} r_2^{13} r_3^{7} c_1^{15} c_2^{16} - r_1^{11} r_2^{13} r_3^{7} c_1^{15} c_2^{16} = \mathbf{0}$$

such that $\mathbf{y} - \mathbf{x} = (7, 4, 7, 6, 1, 6) - (7, 4, 6, 7, 2, 5) = (0, 0, 1, -1, -1, 1) = \mathbf{z} \in B$ and $\mathbf{z}$ is a basic move, see Table 4.

Table 4: :Basic Markov Basis Elements for 3 × 2 Contingency Table.

0	0
+1	−1
−1	+1

0	0
−1	+1
+1	−1

+1	−1
0	0
−1	+1

−1	+1
0	0
+1	−1

+1	−1
−1	+1
0	0

−1	+1
+1	−1
0	0

Then, the toric ideal $I_A =< \{f \in k[\mathbf{u}] : \pi_A(f) = 0\} >$ is the ideal created by binomials $f = \mathbf{u}^{\mathbf{y}} - \mathbf{u}^{\mathbf{x}}, \forall \mathbf{x}, \mathbf{y} \in F_{\mathbf{t}}$ such that $\mathbf{y} - \mathbf{x} = \mathbf{z} \in B \subset ker_{\mathbb{Z}} A$ and this is the relationship between moves and binomials of a toric ideal. To construct a connected graph so as to enable movement all over $F_{\mathbf{t}}$, let $\mathbf{x}$ and $\mathbf{y} \in F_{\mathbf{t}}$. Then, a move $\mathbf{z} = \mathbf{y} - \mathbf{x}$ connects these two elements. Moreover, all fiber elements can be connected by B and forms a connected graph $G_{\mathbf{t},B}$ for given $\mathbf{t}$, where $F_{\mathbf{t}}$ is a set of vertices and B is a set of edges, see Figure 5.

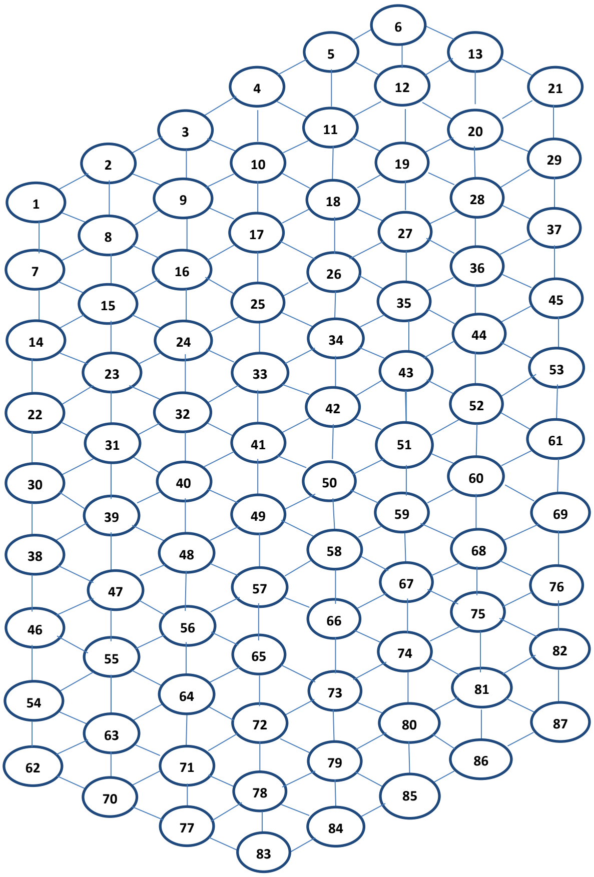


Figure 5: A connected graph $G_{(t,B)}$ for given t.

5. Conclusion

In this study, a new algebraic method has been described to find the set of all independence models of $I \times J$ two-way contingency tables with the same row sums and column sums which called fiber in algebraic statistics. The method has been applied on 3×2 two-way contingency table to show the effectiveness and easily implemented of the purposed method. Furthermore, another advantage of this method is that it only depends on the dimension of the contingency table, the row sums and column sums. However, we cannot walk around all over every fiber that is found by our method, this is the disadvantage of the method. Therefore, we found Markov basis by toric ideal to be able to move all over the fiber.

References

Agnarsson, G. and Greenlaw, R. (2006). *Graph Theory: Modeling, applications, and algorithms.* Prentice-Hall, Inc.

Agresti, A. (1996). *An introduction to categorical data analysis*, volume 135. Wiley New York.

Agresti, A. and Kateri, M. (2011). *Categorical data analysis.* Springer.

Aoki, S., Hara, H., and Takemura, A. (2012). *Markov bases in algebraic statistics*, volume 199. Springer Science & Business Media.

Aoki, S. and Takemura, A. (2008). The largest group of invariance for markov bases and toric ideals. *Journal of Symbolic Computation*, 43(5):342–358.

Diaconis, P., Sturmfels, B., et al. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26(1):363–397.

Dobra, A. and Fienberg, S. E. (2001). Bounds for cell entries in contingency tables induced by fixed marginal totals with applications to disclosure limitation. *Statistical Journal of the United Nations Economic Commission for Europe*, 18(4):363–371.

Dobra, A. and Fienberg, S. E. (2003). Bounding entries in multi-way contingency tables given a set of marginal totals. In *Foundations of Statistical Inference*, pages 3–16. Springer.

Dobra, A., Fienberg, S. E., Rinaldo, A., Slavkovic, A., and Zhou, Y. (2009). Algebraic statistics and contingency table problems: Log-linear models, likelihood estimation, and disclosure limitation. In *Emerging applications of algebraic geometry*, pages 63–88. Springer.

Drton, M., Sturmfels, B., and Sullivant, S. (2008). *Lectures on algebraic statistics*, volume 39. Springer Science & Business Media.

Hosten, S., Khetan, A., and Sturmfels, B. (2005). Solving the likelihood equations. *Foundations of Computational Mathematics*, 5(4):389–407.

Muchiri, J. M., O. G. A. K. N. J. and Kiriungi, L. (2013). Agriculture teachers perception of secondary school agriculture: A case of meru central district, kenya. *Middle-East Journal of Scientific Research*, 17(4):534–538.

Pachter, L. and Sturmfels, B. (2004). Parametric inference for biological sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46):16138–16143.

Pachter, L. and Sturmfels, B. (2005). *Algebraic statistics for computational biology*, volume 13. Cambridge university press.

Pistone, G., Riccomagno, E., and Wynn, H. P. (2000). *Algebraic statistics: Computational commutative algebra in statistics.* CRC Press.

Pistone, G. and Wynn, H. P. (1996). Generalised confounding with gröbner bases. *Biometrika*, 83(3):653–666.

Slavkovic, A. B. and Fienberg, S. E. (2004). Bounds for cell entries in two-way tables given conditional relative frequencies. In *International Workshop on Privacy in Statistical Databases*, pages 30–43. Springer.

Sturmfels, B. (1996). *Gröbner bases and convex polytopes*, volume 8. American Mathematical Soc.

Takemura, A. and Aoki, S. (2004). Some characterizations of minimal markov basis for sampling from discrete conditional distributions. *Annals of the Institute of Statistical Mathematics*, 56(1):1–17.