On connectivity of fibers with positive marginals in multiple logistic regression

Hisayuki Hara*, Akimichi Takemura†‡ and Ruriko Yoshida§

October 2008

Abstract

In this paper we consider exact tests of a multiple logistic regression, where the levels of covariates are equally spaced, via Markov beses. In usual application of multiple logistic regression, the sample size is positive for each combination of levels of the covariates. In this case we do not need a whole Markov basis, which guarantees connectivity of all fibers. We first give an explicit Markov basis for multiple Poisson regression. By the Lawrence lifting of this basis, in the case of bivariate logistic regression, we show a simple subset of the Markov basis which connects all fibers with a positive sample size for each combination of levels of covariates.

Keywords : contingency tables, exact test, Lawrence lifting, Markov bases, MCMC, Segre product

1 Introduction

Diaconis and Sturmfels [1998] developed an algorithm for sampling from conditional distributions for a statistical model of discrete exponential families, based on the algebraic theory of toric ideals. This algorithm is applied to categorical data analysis through the notion of Markov bases. However, often Markov bases are large and difficult to compute. One reason for their large size is that they guarantee connectivity of all fibers (conditional sample spaces). With a given particular data set, on the other hand, we are naturally interested in the connectivity of a particular fiber. However obtaining a subset of a Markov basis for connecting a particular fiber is also a difficult problem in general [Chen et al., 2008]. This problem was already discussed in Section 3 of Diaconis and Sturmfels [1998] concerning "corner minors". In Aoki and Takemura [2005] the case of two-way incomplete tables was studied.

*Department of Technology Management for Innovation, University of Tokyo
†Graduate School of Information Science and Technology, University of Tokyo
‡CREST, JST
§Department of Statistics, University of Kentucky

In most applications of the logistic regression model, for each combination of covariates, the number of successes and the number of failures are observed. The number of trials (i.e. the sum of numbers of successes and failures) for each combination of covariates is usually fixed by a sampling scheme and positive. We call this marginal *the response variable marginal*. Therefore we are usually interested in the connectivity of fibers with positive response variable marginals rather than all the fibers. First, in this paper, we show an explicit form of a Markov basis for multiple Poisson regression. Then, extending the result of Chen et al. [2005], we show an explicit form of a subset of Markov basis, which guarantees the connectivity of every fiber with positive response variable marginals for bivariate logistic regression. We conjecture that a similar subset of Markov basis connects fibers with positive response variable marginals for a general multiple logistic regression. However, it seems difficult to prove this conjecture.

The logistic regression can be understood as the Lawrence lifting of a Poisson regression. Let A denote a configuration defining a toric ideal and let $\Lambda(A)$ denote its Lawrence lifting. Let I_A and $I_{\Lambda(A)}$ denote the respective toric ideals. It is known [Sturmfels, 1996, Theorem 7.1] that the unique minimal Markov basis of $I_{\Lambda(A)}$ coincides with the Graver basis of I_A. Therefore the whole Graver basis of I_A is needed to guarantee the connectivity of all fibers of $\Lambda(A)$. However many of the elements of the Graver basis of I_A seem to be needed to cope with the case of zero response variable marginal frequencies. In Section 4, for the case of bivariate logistic regression, we prove that a smaller Markov basis for the Poisson regression extended to the logistic regression guarantees the connectivity of fibers with positive response variable marginals.

This paper is organized as follows. In Section 2 we summarize results on Markov basis of univariate Poisson regression and results on the connectivity for fibers with positive response variable marginals of univariate logistic regression. In Section 3 we prove a theorem on Markov bases of Segre product of configurations and apply it to multiple Poisson regression. In Section 4 we prove the connectivity of fibers with positive response variable marginals in the case of bivariate logistic regression. Some numerical examples are given in Section 5. We conclude this paper with some discussions in Section 6. Some detailed proofs are in Appendix.

2 Univariate Poisson and logistic regressions

In this section we summarize results on Markov basis of univariate Poisson regression and the connectivity results for fibers with positive response variable marginals of univariate logistic regression. We provide exact statements and detailed proofs of these results, because they are not explicitly given in literature and similar arguments will be repeatedly applied to prove our main theorem in Section 4.

Consider univariate Poisson regression [Diaconis et al., 1998] with the set of levels $\{1, \dots, J\}$ of a covariate. The mean μ_j of independent Poisson random variables X_j, $j = 1, \dots, J$, is modeled as

$$\log \mu_j = \alpha + \beta j, \quad j = 1, \dots, J.$$

The sufficient statistic for the models is $(\sum_{j=1}^J X_j, \sum_{j=1}^J jX_j)$. The first component is the total sample size $n = \sum_{j=1}^J X_j$. The configuration A, i.e. the matrix giving the relation between the observation vector and the sufficient statistic, for this model is given by

$$A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & J \end{pmatrix}. \tag{1}$$

Now we show the minimum-fiber Markov basis [Takemura and Aoki, 2005] for the univariate Poisson regression. The minimum-fiber Markov basis is the union of all minimal Markov bases.

Proposition 1. *Let $\boldsymbol{e}_j$ denote the contingency table with just 1 frequency in the j-th cell. The set of moves*

$$\mathcal{B} = \{\pm(\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_4} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3}) \mid 1 \le j_1 < j_2 \le j_3 < j_4 \le J, \quad j_2 - j_1 = j_4 - j_3\} \tag{2}$$

forms the minimum-fiber Markov basis for the univariate Poisson regression.

Proof. We employ the distance reducing argument of Takemura and Aoki [2005]. Let $x = (x_1, \dots, x_J)$ and $y = (y_1, \dots, y_J)$, $x \neq y$, be two data sets in the same fiber. Let $j_1 = \min\{k \mid x_k \neq y_k\}$. Consider the case $x_{j_1} > y_{j_1}$. The case $x_{j_1} < y_{j_1}$ can be handled by interchanging the roles of x and y. Because the total sample size n is the same in x and y (i.e. $n = \sum_{k=1}^J x_k = \sum_{k=1}^J y_k$), there exists some $j_2 > j_1$ such that $x_{j_2} < y_{j_2}$. Choose the smallest such j_2. Now suppose that $x_k \le y_k$ for all $k \ge j_2$. Then

$$\begin{aligned} 0 = \sum_{k=1}^J k(y_k - x_k) &\ge \sum_{k=j_2}^J j_2(y_k - x_k) - \sum_{k=1}^{j_2-1}(j_2 - 1)(x_k - y_k) \\ &= j_2 \sum_{k=1}^J (y_k - x_k) + \sum_{k=1}^{j_2-1}(x_k - y_k) = \sum_{k=1}^{j_2-1}(x_k - y_k) > 0, \end{aligned}$$

which is a contradiction. Therefore there exists some $j_4 > j_2$, such that $x_{j_4} > y_{j_4}$. Define $j_3 := j_1 + j_4 - j_2$. Then $\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_4} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3}$ can be subtracted from x and the L_1 distance to y becomes smaller. This proves that $\mathcal{B}$ forms a Markov basis.

Now consider a fiber $\mathcal{F}_{2,c}$ with sample size $n = 2$ and a particular value of $c = \sum_{k=1}^k kx_k$. This fiber is written as

$$\mathcal{F}_{2,c} = \{\boldsymbol{e}_j + \boldsymbol{e}'_j \mid 1 \le j \le j' \le J, \ j + j' = c\}$$

and $\mathcal{B}$ consists of all the differences of two elements of these fibers. Since $\mathcal{B}$ forms a Markov basis, every minimal Markov basis needs to connect only these fibers. This proves that $\mathcal{B}$ is the minimum-fiber Markov basis □

We now consider univariate logistic regression [Chen et al., 2005]. Let $\{1, \dots, J\}$ be the set levels of a covariate and let X_{1j} and X_{2j}, $j = 1, \dots, J$, be the numbers of successes and failures, respectively. The probability for success p_j is modeled as

$$\text{logit}(p_j) = \log \frac{p_j}{1 - p_j} = \alpha + \beta j, \qquad j = 1, \dots, J.$$

The sufficient statistics for the model is $(X_{1+}, X_{+1}, \dots, X_{+J}, \sum_{j=1}^{J} jX_{+j})$. Hence moves $z = (z_{ij})$ for the model satisfy $(z_{1+}, z_{+1}, \dots, z_{+J}) = 0$ and

$$\sum_{j=1}^{J} jz_{+j} = 0. \tag{3}$$

The configuration for this model is the Lawrence lifting $\Lambda(A)$ of A in (1):

$$\Lambda(A) = \begin{pmatrix} A & 0 \\ E_J & E_J \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & J \end{pmatrix}, \tag{4}$$

where E_J denotes the $J \times J$ identity matrix.

In general Markov bases of $\Lambda(A)$ become very complicated. In usual applications of the logistic regression model, however, $X_{+j} := X_{1j} + X_{2j}$ is fixed by a sampling scheme and positive. Chen et al. [2005] showed that a simple subset of Markov bases of $\Lambda(A)$ guarantees the connectivity of all fibers satisfying $(X_{+1}, \dots, X_{+J}) > 0$, where the inequality "$> 0$" means that every element is positive.

Let $\boldsymbol{e}_j$ be redefined by a $2 \times J$ integer array with 1 in the $(1, j)$-cell and -1 in the $(2, j)$-cell. Then we can show that the set of moves in (2) connects all fibers with $(X_{+1}, \dots, X_{+J}) > 0$. More strongly, the set of moves is norm-reducing [Takemura and Aoki, 2005] for any two tables x, y in any fiber with positive marginals, i.e. we can make the L_1 distance between x and y smaller by a move from the set.

Proposition 2. *The set of moves*

$$\mathcal{B}_{\Lambda(A)} = \{\pm(\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_4} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3}) \mid 1 \le j_1 < j_2 \le j_3 < j_4 \le J, \quad j_2 - j_1 = j_4 - j_3\} \tag{5}$$

is norm-reducing for all fibers with $(X_{+1}, \dots, X_{+J}) > 0$ for the univariate logistic regression model.

To prove this proposition, we present a simple lemma.

Lemma 1. *Let $z = \{z_{ij}\}$ be any move for the univariate logistic regression. Then there exist $j_1 < j_2$ and $j_3 < j_4$ satisfying the following conditions.*

(a) $z_{1j_1} > 0$, $z_{1j_2} < 0$, $z_{1j_3} < 0$, $z_{1j_4} > 0$;

(b) $z_{1j_1} = 1$ *implies* $j_1 \neq j_4$;

(c) $z_{1j_2} = -1$ *implies* $j_2 \neq j_3$;

(d) $z_{1j} = 0$ *for* $j_1 < j < j_2$ *and* $j_3 < j < j_4$.

Proof. (a), (b) and (c) are obvious from the constraint (3) and $z_{1+} = 0$. We can assume without loss of generality that there exist $j_1 < j_2$ such that $z_{1j_1} > 0$, $z_{1j_2} < 0$, $z_{1j} \ge 0$ for $1 \le j < j_2$ and $z_{1j} = 0$ for $j_1 < j < j_2$. Since there exist $j_2 \le j_3 < j_4$ satisfying (a), (b) and (c), we can choose j_3 and j_4 to satisfy (d). □

We now give a proof of Proposition 2.

Proof of Propsition 2. We employ the distance reducing argument of Takemura and Aoki [2005]. Let x and y be two tables in the same fiber. Then $z := x - y$ is a move. We can assume without loss of generality that there exist $j_1 < j_2 \le j_3 < j_4$ which satisfy the conditions of Lemma 1 and $j_2 - j_1 \le j_4 - j_3$. Define j_5 as $j_5 := j_4 - (j_2 - j_1)$. Then by applying a move

$$z' := -\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_2} + \boldsymbol{e}_{j_5} - \boldsymbol{e}_{j_4},$$

we can reduce the L_1 distance between x and y, because at least one of the following operations can be performed to x or y:

$$\begin{array}{c} \\ i=1 \\ i=2 \end{array}\begin{array}{c} \begin{array}{cccc} j_1 & j_2 & j_5 & j_4 \end{array} \\ \boxed{\begin{array}{rrrr} + & 0+ & 0+ & + \\ 0+ & + & + & 0+ \end{array}} \end{array} + \boxed{\begin{array}{rrrr} -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{array}} = \boxed{\begin{array}{rrrr} 0+ & + & + & 0+ \\ + & 0+ & 0+ & + \end{array}}$$

$$\begin{array}{c} \\ i=1 \\ i=2 \end{array}\begin{array}{c} \begin{array}{cccc} j_1 & j_2 & j_5 & j_4 \end{array} \\ \boxed{\begin{array}{rrrr} 0+ & + & + & 0+ \\ + & 0+ & 0+ & + \end{array}} \end{array} + \boxed{\begin{array}{rrrr} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{array}} = \boxed{\begin{array}{rrrr} + & 0+ & 0+ & + \\ 0+ & + & + & 0+ \end{array}}$$

where $0+$ denote that the cell frequency is nonnegative. $\square$

Chen et al. [2005] introduced a subset of $\mathcal{B}$ which still connects all fibers with $X_{+j} > 0, \forall j$. Chen et al. [2005] did not give a proof of the following theorem.

Theorem 1 (Chen et al. [2005])**.** *The set of moves*

$$\mathcal{B}_0 = \{z \in \mathcal{B} \mid j_2 = j_1 + 1, j_3 = j_4 - 1\} \tag{6}$$

connects every fiber satisfying $(X_{+1}, \dots, X_{+J}) > 0$ *for the univariate logistic regression model.*

Proof. It suffices to show that any move in $z \in \mathcal{B}_{\Lambda(A)}$ of Proposition 2 can be replaced by a series of moves in $\mathcal{B}_0$. To prove this it suffices to show that the L_1 norm of any move $z \in \mathcal{B}_{\Lambda(A)}$, i.e., the L_1 distance between the positive part $x = z^+$ and the negative part $y = z^-$ of $z \in \mathcal{B}_{\Lambda(A)}$, is reduced by moves in $\mathcal{B}_0$. Denote $z := \boldsymbol{e}_{j_1} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3} + \boldsymbol{e}_{j_4}$. We can assume without loss of generality that $j_1 < j_2 \le j_3 < j_4$. We prove it by the induction on $\delta := j_2 - j_1 = j_4 - j_3 \ge 2$.

When $(x_{1,j_1+1}, x_{1,j_4-1}) > 0$ or $(x_{2,j_1+1}, x_{2,j_4-1}) > 0$, we can apply $z' := -\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_1+1} + \boldsymbol{e}_{j_4-1} - \boldsymbol{e}_{j_4}$ to z and

$$z + z' := \boldsymbol{e}_{j_1+1} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3} + \boldsymbol{e}_{j_4-1}$$

as seen from the picture below, where $z_{ij} = 0^*$ denotes that $x_{ij} = y_{ij} > 0$.

$$\begin{array}{c|cccc|} & j_1 & j_1+1 & j_4-1 & j_4 \\ \cline{2-5} i=1 & 1 & 0^* & 0^* & 1 \\ i=2 & -1 & 0 & 0 & -1 \\ \cline{2-5} \end{array} + \begin{array}{|cccc|} \hline -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ \hline \end{array} = \begin{array}{|cccc|} \hline 0 & 1 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ \hline \end{array}$$

$$\begin{array}{c|cccc|} & j_1 & j_1+1 & j_4-1 & j_4 \\ \cline{2-5} i=1 & 1 & 0 & 0 & 1 \\ i=2 & -1 & 0^* & 0^* & -1 \\ \cline{2-5} \end{array} + \begin{array}{|cccc|} \hline -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ \hline \end{array} = \begin{array}{|cccc|} \hline 0 & 1 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ \hline \end{array}$$

From the inductive assumption, we can reduce the L_1 norm of $z + z'$ by moves in $\mathcal{B}_0$.

When $(x_{1,j_1+1}, x_{2,j_3+1}) > 0$ or $(x_{2,j_1+1}, x_{1,j_3+1}) > 0$, we can apply $z' := -\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_1+1} + \boldsymbol{e}_{j_3} - \boldsymbol{e}_{j_3+1}$ to z and

$$z + z' := \boldsymbol{e}_{j_1+1} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3+1} + \boldsymbol{e}_{j_4}$$

as seen from the picture below.

$$\begin{array}{c|cccc|} & j_1 & j_1+1 & j_3 & j_3+1 \\ \cline{2-5} i=1 & 1 & 0^* & -1 & 0 \\ i=2 & -1 & 0 & 1 & 0^* \\ \cline{2-5} \end{array} + \begin{array}{|cccc|} \hline -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ \hline \end{array} = \begin{array}{|cccc|} \hline 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ \hline \end{array}$$

$$\begin{array}{c|cccc|} & j_1 & j_1+1 & j_3 & j_3+1 \\ \cline{2-5} i=1 & 1 & 0 & -1 & 0^* \\ i=2 & -1 & 0^* & 1 & 0 \\ \cline{2-5} \end{array} + \begin{array}{|cccc|} \hline -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ \hline \end{array} = \begin{array}{|cccc|} \hline 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ \hline \end{array}$$

From the inductive assumption, we can reduce the L_1 norm of $z + z'$ by moves in $\mathcal{B}_0$. In the case that $(x_{1,j_2-1}, x_{2,i_4-1}) > 0$ or $(x_{2,j_2-1}, x_{1,i_4-1}) > 0$, the proof is similar.

Suppose that

$$x_{2,j_1+1} = x_{1,j_2-1} = x_{2,j_3+1} = x_{1,j_4-1} = 0.$$

We note that this implies that

$$x_{1,j_1+1} > 0, \quad x_{2,j_2-1} > 0, \quad x_{1,j_3+1} > 0, \quad x_{2,j_4-1} > 0.$$

Then there exists $j_3 < j_5 < j_4$ such that

$$x_{1j_5} > 0, \quad x_{2,j_5+1} > 0$$

as in the following picture.

$$\begin{array}{c|cccccccccc|} & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \cline{2-11} i=1 & 1 & 0^* & 0 & -1 & -1 & 0^* & 0^* & 0 & 0 & 1 \\ i=2 & -1 & 0 & 0^* & 1 & 1 & 0 & 0 & 0^* & 0^* & -1 \\ \cline{2-11} \end{array}$$

By applying

$$z' := -\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_1+1} + \boldsymbol{e}_{j_5} - \boldsymbol{e}_{j_5+1}$$

and

$$z'' := -\boldsymbol{e}_{j_5} + \boldsymbol{e}_{j_5+1} + \boldsymbol{e}_{j_4-1} - \boldsymbol{e}_{j_4}$$

to z in this order, we obtain

$$z + z' + z'' = \boldsymbol{e}_{j_1+1} - \boldsymbol{e}_{j_2} - \boldsymbol{e}_{j_3} - \boldsymbol{e}_{j_4-1}$$

as in the following picture.

$$\begin{array}{r|cccccccccc|}
 & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \hline
i=1 & 1 & 0^* & 0 & -1 & -1 & 0^* & 0^* & 0 & 0 & 1 \\
i=2 & -1 & 0 & 0^* & 1 & 1 & 0 & 0 & 0^* & 0^* & -1 \\ \hline
\end{array}$$

$$+\ \begin{array}{r|cccccccccc|}
 & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \hline
i=1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
i=2 & 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ \hline
\end{array}$$

$$=\ \begin{array}{r|cccccccccc|}
 & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \hline
i=1 & 0 & 1 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & -1 \\
i=2 & 0 & -1 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 1 \\ \hline
\end{array}$$

$$\begin{array}{r|cccccccccc|}
 & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \hline
i=1 & 0 & 1 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & -1 \\
i=2 & 0 & -1 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 1 \\ \hline
\end{array}$$

$$+\ \begin{array}{r|cccccccccc|}
 & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \hline
i=1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 \\
i=2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ \hline
\end{array}$$

$$=\ \begin{array}{r|cccccccccc|}
 & j_1 & j_1+1 & j_2-1 & j_2 & j_3 & j_3+1 & j_5 & j_5+1 & j_4-1 & j_4 \\ \hline
i=1 & 0 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\
i=2 & 0 & -1 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 \\ \hline
\end{array}$$

From the inductive assumption, we can reduce the L_1 norm of $z + z'$ by moves in $\mathcal{B}_0$. In the case that $x_{1,j_1+1} = x_{2,j_2-1} = x_{1,j_3+1} = x_{2,j_4-1} = 0$, the proof is similar. $\square$

3 Markov bases for models of Segre product type

In the previous section we considered univariate Poisson regression and logistic regression. We now consider generalizing the results to multiple regression. In this section we show an explicit form of Markov basis for multiple Poisson regression. Therefore an extension of Proposition 1 to multiple regression is straightforward. In contrast, as we see in the next section, it is difficult to generalize the results of univariate logistic regression to multiple logistic regression.

Multiple Poisson regression is a Segre product of univariate Poisson regressions. Quadratic Gröbner bases of Segre products is already discussed in Aoki et al. [2008]. However Theorem 2 below is about Markov bases (rather than Gröbner bases) and it is applicable even if the component configurations do not possess quadratic Gröbner bases.

Consider two configurations $A = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_J)$ and $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_K)$, where $\boldsymbol{a}_j$ and $\boldsymbol{b}_k$ are column vectors. We assume the homogeneity, i.e., there exist weight vectors w, v such that $\langle w, \boldsymbol{a}_j \rangle = 1,\ \forall j,\ \langle v, \boldsymbol{b}_k \rangle = 1,\ \forall k$. The configuration $A \otimes B$ of the Segre product of A and B is defined as

$$A \otimes B = \Big(\boldsymbol{a}_j \oplus \boldsymbol{b}_k,\ j = 1, \dots, J, k = 1, \dots, K\Big), \qquad \boldsymbol{a}_j \oplus \boldsymbol{b}_k = \begin{pmatrix} \boldsymbol{a}_j \\ \boldsymbol{b}_k \end{pmatrix}.$$

If both A and B are configurations of the form (1) for the univariate Poisson regression model, then $A \otimes B$ corresponds to the bivariate Poisson regression model, where X_{jk} is independently distributed according to Poisson distribution with mean μ_{jk}, which is modeled as

$$\log \mu_{jk} = \mu + \alpha j + \beta k, \qquad j = 1, \dots, J,\ k = 1, \dots, K.$$

Let $X = (X_{jk})_{j=1,\dots,J,k=1,\dots,K}$ denote a table of observed frequencies. The sufficient statistic for the Segre product $A \otimes B$ is given by

$$\sum_j \boldsymbol{a}_j X_{j+}, \quad \sum_k \boldsymbol{b}_k X_{+k}.$$

Therefore $z = (z_{jk})$ is a move for $A \otimes B$ if and only if

$$0 = \sum_j \boldsymbol{a}_j z_{j+}, \quad 0 = \sum_k \boldsymbol{b}_k z_{+k}. \tag{7}$$

Given Markov bases $\mathcal{B}_A$ and $\mathcal{B}_B$ for A and B, respectively, our goal is to construct a Markov basis for the Segre product $A \otimes B$. Denote the elements of $\mathcal{B}_A$ by $z^A = (z_1^A, \dots, z_J^A)$. Let $z_j^{A,+} = \max(z_j^A, 0)$ be the positive part and $z_j^{A,-} = \max(-z_j^A, 0)$ be the negative part of z_j^A. Let $\deg z^A = \sum_{j=1}^J z_j^{A,+} = \sum_{j=1}^J z_j^{A,-}$ be the degree of z^A. Now z^A is uniquely written as

$$z^A = \sum_{h=1}^{\deg z^A} (\boldsymbol{e}_{j_h} - \boldsymbol{e}_{j'_h}),$$

where $j_1 \le \dots \le j_{\deg z^A}$ and $j'_1 \le \dots \le j'_{\deg z^A}$. Let $\boldsymbol{e}_{jk}$ denote a $J \times K$ table with 1 at the cell (j, k) and 0 everywhere else. Now choose arbitrary $1 \le k_1, \dots, k_{\deg z^A} \le K$ and define

$$z^A(k_1, \dots, k_{\deg z^A}) = \sum_{h=1}^{\deg z^A} (\boldsymbol{e}_{j_h k_h} - \boldsymbol{e}_{j'_h k_h}).$$

We call $z^A(k_1, \dots, k_{\deg z^A})$ a "distribution" of z^A by coordinates $k_1, \dots, k_{\deg z^A}$. Note that $k_1, \dots, k_{\deg z^A}$ are not ordered. Similarly define the distribution $z^B(j_1, \dots, j_{\deg z^B})$ of a move $z^B \in \mathcal{B}_B$.

In addition to these moves we also consider the basic moves $z(j_1, j_2; k_1, k_2) = \boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_2k_2} - \boldsymbol{e}_{j_1k_2} - \boldsymbol{e}_{j_2k_1}$ of the form $\begin{array}{cc} +1 & -1 \\ -1 & +1 \end{array}$. We now have the following theorem.

Theorem 2. *The set of basic moves and the set of moves of the form* $z^A(k_1, \ldots, k_{\deg z^A})$, $1 \leq k_1, \ldots, k_{\deg z^A} \leq K$, $z^A \in \mathcal{B}_A$, $z^B(j_1, \ldots, j_{\deg z^B})$, $1 \leq j_1, \ldots, j_{\deg z^B} \leq J$, $z^B \in \mathcal{B}_B$, *form a Markov basis for the Segre product* $A \otimes B$.

A proof of this theorem is given in Appendix. In Theorem 2 we have considered Segre product of two configurations. By a recursive argument, a Markov basis for the Segre product of arbitrary number of configurations $A_1 \otimes \cdots \otimes A_m$ is given as follows. Let $\mathcal{B}_{A_j}$ be a Markov basis for the configuration A_j, $j = 1, \ldots, m$. Write $[J] = \{1, \ldots, J\}$ and let

$$\bar{\mathcal{J}}_j = [J_1] \times \cdots \times [J_{j-1}] \times [J_{j+1}] \times \cdots \times [J_m].$$

Let $z^{A_j} \in \mathcal{B}_{A_j}$ and let $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{\deg z^{A_j}} \in \bar{\mathcal{J}}_j$. Now define

$$z^{A_j}(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{\deg z^{A_j}}) = \sum_{h=1}^{\deg z^{A_j}} (\boldsymbol{e}_{j_h \boldsymbol{k}_h} - \boldsymbol{e}_{j'_h \boldsymbol{k}_h}),$$

where $\boldsymbol{e}_{j,\boldsymbol{k}}$ is an m-way table with 1 at the cell $(j, \boldsymbol{k})$ and 0 everywhere else. Then we have the following corollary to Theorem 2.

Corollary 1. *The set of square-free degree two moves for the complete independence model of* $J_1 \times \cdots \times J_m$ *contingency tables and the set of moves of the form* $z^{A_j}(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{\deg z^A})$, $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{\deg z^A} \in \bar{\mathcal{J}}_j$, $j = 1, \ldots, m$, *form a Markov basis for the Segre product* $A_1 \otimes \cdots \otimes A_m$.

Minimality of the Markov basis constructed in Theorem 2 is not clear at the present. However the maximum degree of moves in the Markov basis for $A_1 \otimes \cdots \otimes A_m$ is bounded by the maximum degree of moves in $\mathcal{B}_{A_1}, \ldots \mathcal{B}_{A_m}$.

4 Connectivity of fibers of positive marginals in bivariate logistic regression

In this section we consider the extension of the results in univariate logistic regression model to bivariate logistic regression model. Let $\{1, \ldots, J\}$ and $\{1, \ldots, K\}$ be the sets levels of two covariates. Let X_{1jk} and X_{2jk}, $j = 1, \ldots, J$, $k = 1, \ldots, K$, be the numbers of successes and failures, respectively, for level (j, k). The probability for success p_{1jk} is modeled as

$$\mathrm{logit}(p_{1jk}) = \log\left(\frac{p_{1jk}}{1 - p_{1jk}}\right) = \mu + \alpha j + \beta k, \tag{8}$$
$$j = 1, \ldots, J, \quad k = 1, \ldots, K.$$

The sufficient statistics for this model is $X_{1++},\ \sum_{j=1}^{J} jX_{1j+},\ \sum_{k=1}^{K} kX_{1+k},\ X_{+jk},\ \forall j,k$. Hence moves $Z=(z_{ijk})$ for the model satisfy

$$z_{1++}=0,\quad \sum_{j=1}^{J} jz_{1j+}=0,\quad \sum_{k=1}^{K} kz_{1+k}=0,\quad z_{+jk}=0,\ \forall j,k.$$

Let

$$B=\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & K \end{pmatrix}.$$

and let A be defined as in (4). Then the configuration for the bivariate logistic regression model is the Lawrence lifting of Segre product $\Lambda(A\otimes B)$. Here we consider a set of moves which connects every fiber satisfying $X_{+jk}>0,\ \forall j,k$.

Definition 1. *Let $\boldsymbol{e}_{jk}=(e_{ijk})$ be redefined as an integer array with 1 at the cell $(1jk)$, -1 at the cell $(2jk)$ and 0 everywhere else. Define $\mathcal{B}_{\Lambda(A\otimes B)}$ as the set of moves $z=(z_{ijk})$ satisfying the following conditions,*

1. $z=\boldsymbol{e}_{j_1k_1}-\boldsymbol{e}_{j_2k_2}-\boldsymbol{e}_{j_3k_3}+\boldsymbol{e}_{j_4k_4}$;
2. $(j_1,k_1)-(j_2,k_2)=(j_3,k_3)-(j_4,k_4)$.

$\mathcal{B}_{\Lambda(A\otimes B)}$ is an extension of $\mathcal{B}_{\Lambda(A)}$ in Proposition 2 to the bivariate model (8). We note that the $(i=1)$-slice of a moves (z_{1jk}) in $\mathcal{B}_{\Lambda(A\otimes B)}$ is a move of the Markov basis defined in Theorem 2. Now we present the main theorem of this paper.

Theorem 3. *$\mathcal{B}_{\Lambda(A\otimes B)}$ connects every fiber satisfying $X_{+jk}>0,\ \forall j,k$.*

A proof of this theorem is given in Appendix. We give some examples of moves in $\mathcal{B}_{\Lambda(A\otimes B)}$.

(1) $k_1=\cdots=k_4$

$$\begin{array}{c|cccc|} & j_1 & j_2 & j_3 & j_4 \\ \hline k_1 & 1 & -1 & -1 & 1 \\ \hline \end{array}$$

(2) $k_1=\cdots=k_4$ and $j_2=j_3$

$$\begin{array}{c|ccc|} & j_1 & j_2 & j_4 \\ \hline k_1 & 1 & -2 & 1 \\ \hline \end{array}$$

(3) $k_1=k_2$ $(k_3=k_4)$

$$\begin{array}{c|cccc|} & j_1 & j_2 & j_3 & j_4 \\ \hline k_1 & 1 & -1 & 0 & 0 \\ k_3 & 0 & 0 & -1 & 1 \\ \hline \end{array}$$

(4) $k_1=k_2$ and $j_2=j_3$

$$\begin{array}{c|ccc|} & j_1 & j_2 & j_4 \\ \hline k_1 & 1 & -1 & 0 \\ k_3 & 0 & -1 & 1 \\ \hline \end{array}$$

(5) $(j_2,k_2)=(j_3,k_3)$

$$\begin{array}{c|ccc|} & j_1 & j_2 & j_4 \\ \hline k_1 & 1 & 0 & 0 \\ k_2 & 0 & -2 & 0 \\ k_4 & 0 & 0 & 1 \\ \hline \end{array}$$

(6) $k_1=k_4$ and $j_2=j_3$

$$\begin{array}{c|ccc|} & j_1 & j_2 & j_4 \\ \hline k_2 & 0 & -1 & 0 \\ k_1 & 1 & 0 & 1 \\ k_3 & 0 & -1 & 0 \\ \hline \end{array}$$

5 Numerical examples

5.1 Data on coronary heart disease incidence

Table 1 refers to coronary heart disease incidence in Framingham, Massachusetts [Cornfield, 1962, Agresti, 1990]. A sample of male residents, aged 40 through 50, were classified on

Table 1: Data on coronary heart disease incidence

	Blood Pressure	Serum Cholesterol (mg/100ml) 1 < 200	2 200-209	3 210-219	4 220-244	5 245-259	6 260-284	7 > 284
1	< 117	2/53	0/21	0/15	0/20	0/14	1/22	0/11
2	117-126	0/66	2/27	1/25	8/69	0/24	5/22	1/19
3	127-136	2/59	0/34	2/21	2/83	0/33	2/26	4/28
4	137-146	1/65	0/19	0/26	6/81	3/23	2/34	4/23
5	147-156	2/37	0/16	0/6	3/29	2/19	4/16	1/16
6	157-166	1/13	0/10	0/11	1/15	0/11	2/13	4/12
7	167-186	3/21	0/5	0/11	2/27	2/5	6/16	3/14
8	> 186	1/5	0/1	3/6	1/10	1/7	1/7	1/7

Source : Cornfield [1962]

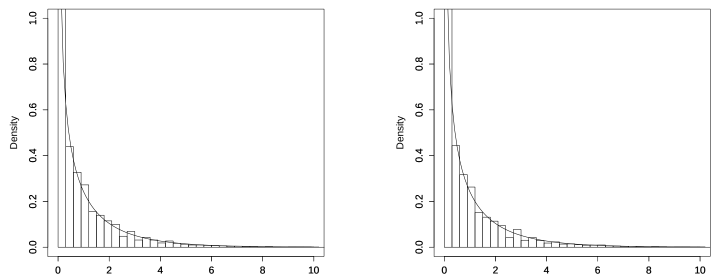
blood pressure and serum cholesterol concentration. 2/53 in the (1,1) cell means that there are 53 cases, of whom 2 exhibited heart disease. We examine the goodness-of-fit of the model (8) with $J = 7$ and $K = 8$. We first test the null hypotheses $\mathrm{H}_\alpha : \alpha = 0$ and $\mathrm{H}_\beta : \beta = 0$ versus (8) using the likelihood ratio statistics L_α and L_β. Then we have $L_\alpha = 18.09$ and $L_\beta = 22.56$ and the asymptotic p-values are 2.107×10^{-5} and 2.037×10^{-6}, respectively, from the asymptotic distribution χ^2_1. We computed the exact distribution of L_α and L_β via Monte Carlo Markov chain (MCMC) with the sets of moves $\mathcal{B}_{\Gamma(A)}$ and $\mathcal{B}_0$ discussed in Section 2. See the last paragraph of Section 6 on sampling under H_α and H_β. In all experiments in this paper, we sampled 100,000 tables after 50,000 burn-in steps. Figure 1 and 2 represent histograms of L_α and L_β. The solid lines in the figures represent the density function of the asymptotic distribution χ^2_1. The estimated p-values are 1.0×10^{-6} for all cases. Therefore both H_α and H_β are rejected. We can see from the figures that there are little differences between two histograms computed with $\mathcal{B}_{\Gamma(A)}$ and $\mathcal{B}_0$.

Next we set (8) as a null hypothesis and test it versus the following ANOVA type logit model,

$$\mathrm{H}_1 : \mathrm{logit}(p_{1jk}) = \log\left(\frac{p_{1jk}}{1 - p_{1jk}}\right) = \mu + \alpha_j + \beta_k, \tag{9}$$

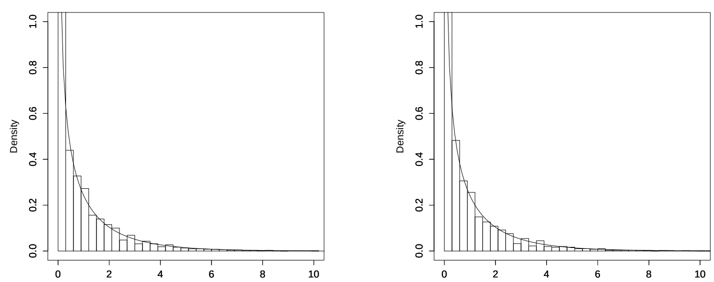
where $\sum_{j=1}^{J} \alpha_j = 0$ and $\sum_{k=1}^{K} \beta_k = 0$ by likelihood ratio statistic L_0. The value of L_0 is 13.07587 and the asymptotic p-value is 0.2884 from the asymptotic distribution χ^2_{11}. We computed the exact distribution of L_0 via MCMC with $\mathcal{B}_{\Gamma(A \otimes B)}$ defined in Definition 1. As an extension of $\mathcal{B}_0$ in Theorem 1 to the bivariate model (8), we define $\mathcal{B}_0^2$ by the set of moves $z = \boldsymbol{e}_{j_1k_1} - \boldsymbol{e}_{j_2k_2} - \boldsymbol{e}_{j_3k_3} + \boldsymbol{e}_{j_4k_4}$ satisfying $(j_1, k_1) - (j_2, k_2) = (j_3, k_3) - (j_4, k_4)$ is either of $(\pm 1, 0)$, $(0, \pm 1)$, $(\pm 1, \pm 1)$ or $(\pm 1, \mp 1)$. We also computed the exact distribution of L_0 with $\mathcal{B}_0^2$. Figure 3 represents histograms of L_0 computed with $\mathcal{B}_{\Gamma(A \otimes B)}$ and $\mathcal{B}_0^2$. The estimated p-values are 0.2706 with $\mathcal{B}_{\Gamma(A \otimes B)}$ and 0.2958 with $\mathcal{B}_0^2$. Therefore the model (8) is accepted.

The p-values estimated with $\mathcal{B}_{\Gamma(A \otimes B)}$ and $\mathcal{B}_0^2$ are close and there are little differences



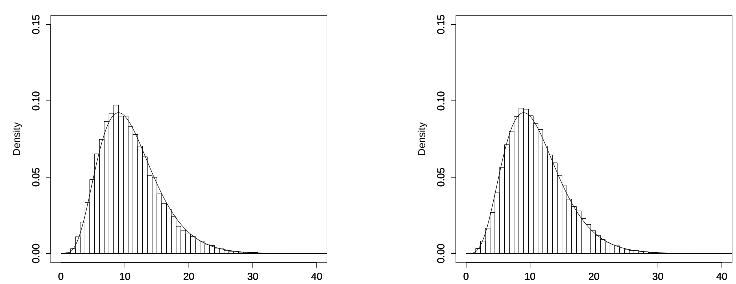
(a) A histogram with $\mathcal{B}_{\Gamma(A)}$ (b) A histogram with $\mathcal{B}_0$

Figure 1: Histograms of L_α via MCMC with $\mathcal{B}_{\Gamma(A)}$ and $\mathcal{B}_0$



(a) A histogram with $\mathcal{B}_{\Gamma(A)}$ (b) A histogram with $\mathcal{B}_0$

Figure 2: Histograms of L_β via MCMC with $\mathcal{B}_{\Gamma(A)}$ and $\mathcal{B}_0$



(a) A histogram with $\mathcal{B}_{\Lambda(A\otimes B)}$ (b) A histogram with $\mathcal{B}_0^2$

Figure 3: Histograms of L_0 via MCMC with $\mathcal{B}_{\Lambda(A\otimes B)}$ and $\mathcal{B}_0^2$

between two histograms. From the results of Theorem 1 and this numerical experiment, $\mathcal{B}_0^2$ is also expected to connect every fiber with positive response variable marginals. However the theoretical proof of it is not clear at the present and is left to our future research.

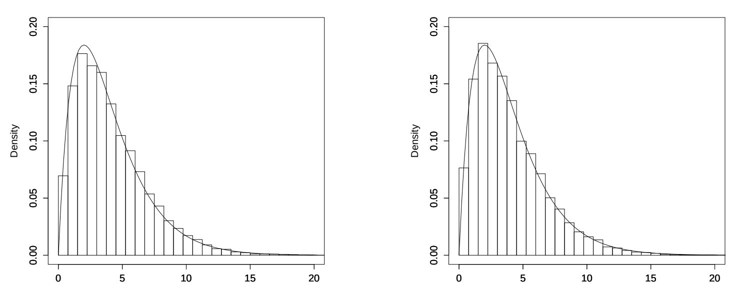
5.2 Data on occurrence of esophageal cancer

Table 2: Data on occurrence of esophageal cancer

		Age					
	Alcohol	1	2	3	4	5	6
	Consumption	25-34	35-44	45-54	55-64	65-74	75+
0	Low	0/106	5/169	21/159	34/173	36/124	8/39
1	High	1/10	4/30	25/54	42/69	19/37	5/5

Source : Breslow and Day [1980]

The second example is from Table 4.16 in Christensen [1997] (data source is from Breslow and Day [1980]). Table 2 refers to the occurrence of esophageal cancer in Frenchmen which were classified on ages and dummy variable on alcohol consumption. We test the goodness-of-fit of the model (8) with $J = 6$ and $K = 2$ by likelihood ratio statistics L_0 via MCMC. Then the value of L_0 is 20.89 and the asymptotic p-value is 0.0003330 from the asymptotic distribution χ^2_4. We computed the exact distribution of L_0 via MCMC with $\mathcal{B}_{\Gamma(A\otimes B)}$ and $\mathcal{B}_0^2$. Figure 4 represents the histograms of L_0. The estimated p-values are 0.00011 with $\mathcal{B}_{\Gamma(A\otimes B)}$ and 0.00055 with $\mathcal{B}_0^2$. Therefore the model (8) is rejected at the significance level of 1%.



(a) a histogram with $\mathcal{B}_{\Lambda(A\otimes B)}$ (b) a histogram with $\mathcal{B}_0^2$

Figure 4: Histograms of L_0 via MCMC with $\mathcal{B}_{\Lambda(A\otimes B)}$ and $\mathcal{B}_0^2$

6 Concluding remarks

In Theorem 3 we showed the connectivity result for bivariate logistic regression. A natural extension of Theorem 3 to m covariates is given as follows. Let $\boldsymbol{j} = (j_1, \ldots, j_m)$ denote the combination of m levels and let $\boldsymbol{e_j}$ denote an array with 1 at the cell $(1, \boldsymbol{j})$ and -1 at the cell $(2, \boldsymbol{j})$. Define $\mathcal{B}_{\Lambda(A_1\otimes\cdots\otimes A_m)}$ as the set of the following moves z:

1. $z = \boldsymbol{e}_{\boldsymbol{j}_1} - \boldsymbol{e}_{\boldsymbol{j}_2} - \boldsymbol{e}_{\boldsymbol{j}_3}, + \boldsymbol{e}_{\boldsymbol{j}_4}$

2. $\boldsymbol{j}_1 - \boldsymbol{j}_2 = \boldsymbol{j}_3 - \boldsymbol{j}_4$.

Then we conjecture the following.

Conjecture 1. *The set of moves $\mathcal{B}_{\Lambda(A_1 \otimes \cdots \otimes A_m)}$ connects every fiber with positive response marginals for the logistic regression with m covariates.*

The separation lemma and some steps of the proof of Theorem 3 can be easily generalized to multiple logistic regression. However many steps of our proof, especially those for Cases 3 and 5, are restricted to the two-dimensional case.

As discussed in Section 5 it seems that we can further restrict to the set of moves $z = \boldsymbol{e}_{\boldsymbol{j}_1} - \boldsymbol{e}_{\boldsymbol{j}_2} - \boldsymbol{e}_{\boldsymbol{j}_3} + \boldsymbol{e}_{\boldsymbol{j}_4}$, where the elements of $\boldsymbol{j}_1 - \boldsymbol{j}_2 = \boldsymbol{j}_3 - \boldsymbol{j}_4$ are ± 1 or 0. Hence a stronger conjecture (even for the case of $m = 2$) is given as follows.

Conjecture 2. *The subset of moves from $\mathcal{B}_{\Lambda(A_1 \otimes \cdots \otimes A_m)}$ such that the elements of $\boldsymbol{j}_1 - \boldsymbol{j}_2 = \boldsymbol{j}_3 - \boldsymbol{j}_4$ are ± 1 or 0 connects every fiber with positive response marginals for the logistic regression with m covariates.*

In this paper we considered logistic regression, which is the Lawrence lifting of Poisson regression. Our Theorem 2 describes Markov bases for a general Segre product of configurations. Therefore it is interesting, in practice, to investigate connectivity result for Lawrence lifting of a general Segre product of configurations.

In the bivariate logistic regression, it is interesting to test the null hypothesis that the coefficient of one of the covariates is zero. Generating random samples under the null hypothesis is simple because it reduces to univariate logistic regression as follows. In (8) consider the null hypothesis $H_0 : \beta = 0$. Given observed data (x_{ijk}), we can generate random sample from the null conditional distribution by MCMC procedure for the marginals (x_{1j+}), $j = 1, \ldots, J$. Then for each j, we can sample x_{1jk}, $k = 1, \ldots, K$, by the random sampling without replacement.

Appendix

A Proof of Theorem 2

Let x and y be two tables in the same fiber. Write $z = x - y$. First consider the case that x and y already have the same marginals:

$$0 = z_{j+}, \forall j, \quad 0 = z_{+k}, \forall k.$$

Then, as is well known for two-way complete independence model, we can use the basic moves to move from X to Y. Note that (7) is always satisfied in these steps.

Next consider the case that the row sums are already the same

$$z_{j+} = 0, \ j = 1, \ldots, J,$$

but the column sums are not yet the same. For the moment, ignoring joint frequencies, just look at the column sums of x and y:

$$(x_{+1}, \dots, x_{+K}), \quad (y_{+1}, \dots, y_{+K})$$

We can use the moves of $\mathcal{B}_B$ to move from the marginal frequency $(x_{+1}, \dots, x_{+K})$ to the marginal frequency $(y_{+1}, \dots, y_{+K})$. However, of course we have to worry about the joint frequencies and the row sums. Here the idea is that we can "distribute" moves of $\mathcal{B}_B$ to the cells of the $J \times K$ table, in such a way that we never disturb the row sums. This way, we can make column sums equal, while always keeping the row sums equal. Consider a situation that a move z^B of $\mathcal{B}_B$ can be added to $(x_{+1}, \dots, x_{+K})$. Then we have

$$x_{+k} \geq z_k^{B,-}, \quad k = 1, \dots, K.$$

This shows that in each column k with $c_k = z_k^{B,-} > 0$, there are at least c_k positive frequencies of x, i.e., there exists indices $j_{1,k}, \dots, j_{c_k,k}$ such that

$$(x_{1k}, \dots, x_{Jk}) - (\boldsymbol{e}_{j_{1,k}} + \cdots + \boldsymbol{e}_{j_{c_k,k}}) \geq 0.$$

Here "≥ 0" means that every component of the left-hand side is non-negative. Collect the indices $j_{1,k}, \dots, j_{c_k,k}$ for all k with $z_k^{B,-} > 0$ as $j_1, \dots, j_{\deg z^B}$. Then $z^B(j_1, \dots, j_{\deg z^B})$ can be added to x. Note that z^B is added to the marginal frequencies $(x_{+1}, \dots, x_{+K})$, but the move does not change the row sums of x. This argument implies that the set of moves $z^B(j_1, \dots, j_{\deg z^B})$ are sufficient for connecting two tables with the same row sums.

Lastly we consider the case that neither the row sums nor the column sums are the same for x and y. We can employ a "greedy algorithm", in which we first look at the row sums only and try to make the row sums equal, because the column sums can be adjusted later by the above argument. Now in the above argument, with the roles of the rows and the columns interchanged, we can ignore the fact that the column sums are not yet equal. We can use the same procedure as above. Therefore, by applying a move of the form $z^A(k_1, \dots, k_{\deg z^A})$ we do not change the column sums of x and y. Then we can make the row sums of x and y equal, while not changing the column sums of x and y.

B A separation lemma

Here we prove a lemma, which is needed for our proof of Theorem 3.

Lemma 2. *Let $\mathcal{I} = [J] \times [K]$ and let S^+ and S^- be disjoint subsets of $\mathcal{I}$ satisfying the following properties:*

1. $(j, k) \in S^+,\ j' \leq j,\ k' \leq k \quad \Rightarrow \quad (j', k') \in S^+.$
2. $(j, k) \in S^-,\ j' \geq j,\ k' \geq k \quad \Rightarrow \quad (j', k') \in S^-.$
3. *There are no distinct four points $(j_1, k_1) \in S^+, (j_2, k_2) \in S^-, (j_3, k_3) \notin S^+, (j_4, k_4) \in S^+$ and there are no distinct four points $(j_1, k_1) \in S^-, (j_2, k_2) \in S^+, (j_3, k_3) \notin S^-, (j_4, k_4) \in S^-$ such that*

$$(j_1, k_1) - (j_2, k_2) = (j_3, k_3) - (j_4, k_4).$$

Then there exists a line with rational slope separating S^+ and S^-, i.e. there exist integers a, b, c, $((a, b) \neq (0, 0))$, such that

$$S^+ \subset \{(j, k) \in \mathcal{I} \mid aj + bk \leq c\}, \qquad S^- \subset \{(j, k) \in \mathcal{I} \mid aj + bk \geq c\}. \tag{10}$$

Proof. The lemma obviously holds if S^+ or S^- is empty. Therefore we only need to consider case that S^+ and S^- are non-empty. Define $j_l = \min\{j \mid \exists (j, k) \in S^-\}$ and for $j \in \{j_l, j_l + 1, \dots, J\}$ define $f(j) = \min\{k \mid (j, k) \in S^-\}$. Let f^* be the largest convex minorant [Moriguti, 1953] of $f(j)$, $j \in \{j_l, j_l + 1, \dots, J\}$, i.e. $f^*(\cdot)$ is the maximum among convex functions not exceeding $f(j)$ for each $j \in \{j_l, j_l + 1, \dots, J\}$. Then f^* consists of finite number of line segments. Let $j_1 < j_4$ be endpoints of a line segment and let $L^*_{j_1,j_4}$ denote the line segment. Then $(j_1, f(j_1)), (j_4, f(j_4)) \in S^-$. Also by construction of f^*,

$$(j, k) \in S^-,\ j_1 \leq j \leq j_4 \quad \Rightarrow \quad k \geq f^*(j).$$

Therefore every point strictly below $L^*_{j_1,j_4}$ belongs to S^+. Consider the rectangular region of integer points

$$R_{j_1,j_4} = \{j_1, \dots, j_4\} \times \{f(j_4), \dots, f(j_1)\}.$$

If there exists a point $(j_2, k_2) \in S^+ \cap R_{j_1,j_4}$ strictly above the line segment $L^+_{j_1,j_4}$, let $(j_3, k_3) = (j_1, k_1) - (j_2, k_2) + (j_4, k_4) \in (S^+)^C \cap R_{j_1,j_4}$ and condition 3 of the lemma is violated. This shows that no point of R_{j_1,j_4} strictly above $L^+_{j_1,j_4}$ belongs to S^-. Also the points above R_{j_1,j_4} belong to S^+ by the monotonicity condition (2). Therefore $L^+_{j_1,j_4}$ is a separating line for the interval $\{j_1, \dots, j_4\}$.

Now we similarly construct the smallest concave majorant $f_*(j)$ for S^-. Then by a hyperplane separation theorem for two convex sets, there exists a line with rational slope between $f_*(j)$ and $f^*(j)$. This prove the lemma. □

C Proof of Theorem 3

Let $x := \{x_{ijk}\}$ and $y := \{y_{ijk}\}$ be two $2 \times J \times K$ tables in the same fiber satisfying $x_{+jk} = y_{+jk} > 0$. Then $z := \{z_{ijk}\} = x - y$ is a move for $\Lambda(A \otimes B)$. Let z^1 denote the $(i = 1)$-slice of z. As mentioned in Section 4, z satisfies $z_{+jk} = 0$, $z_{i++} = 0$, $\forall i, j, k$, and

$$\sum_{j=1}^{J} j z_{1j+} = 0, \qquad \sum_{k=1}^{K} k z_{1+k} = 0. \tag{11}$$

Note that $z_{i++} = 0$ implies

$$\sum_{j=1}^{J} j z_{1j+} = 0 \quad \Leftrightarrow \quad \sum_{j=1}^{J} (J - j + 1) z_{1j+} = 0. \tag{12}$$

Similarly $\sum_{k=1}^{K} k z_{1+k} = 0 \ \Leftrightarrow\ \sum_{k=1}^{K} (K - k + 1) z_{1+k} = 0$. This implies that when we consider a sign pattern of a move, we can arbitrarily choose directions for two factors j and k.

Let $\mathcal{I}^+$ and $\mathcal{I}^-$ be the multisets of indices defined by

$$\mathcal{I}^+ := \{(j,k) \mid z_{1jk} > 0\}, \quad \mathcal{I}^- := \{(j,k) \mid z_{1jk} < 0\},$$

where the multiplicity of (j,k) in $\mathcal{I}^+$ and $\mathcal{I}^-$ is $|z_{1jk}|$.

Suppose that $(j_1,k_1) \in \mathcal{I}^+$, $(j_2,k_2) \in \mathcal{I}^-$ and $j_1 < j_2$. Then we note that there exist $j_3 < j_4$, k_3 and k_4 satisfying

$$(j_3,k_3) \in \mathcal{I}^- \setminus \{(j_2,k_2)\}, \quad (j_4,k_4) \in \mathcal{I}^+ \setminus \{(j_1,k_1)\} \tag{13}$$

from (11). If $k_1 < k_2$ and $k_3 > k_4$, there exists $k_5 < k_6$, j_5 and j_6 satisfying

$$(j_5,k_5) \in \mathcal{I}^- \setminus \{(j_2,k_2),(j_3,k_3)\}, \quad (j_6,k_6) \in \mathcal{I}^+ \setminus \{(j_1,k_1),(j_3,k_4)\}.$$

Write $y(j_1,j_2;k_1,k_2) = \boldsymbol{e}_{j_1k_1} - \boldsymbol{e}_{j_2k_2}$. When a move z includes $y(i_1,i_2;j_1,j_2)$, we denote it by $y(i_1,i_2;j_1,j_2) \subset z$.

Case 1. We first consider the case where there exist j_0, j_1, j_2, k_0, k_1 and k_2 such that

$$z_{1j_0k_1} > 0, \quad z_{1j_0k_2} < 0, \tag{14}$$

$$z_{1j_1k_0} > 0, \quad z_{1j_2k_0} < 0. \tag{15}$$

Without loss of generality we assume $j_1 < j_2$ and $k_1 < k_2$. Let $S^+ = \{(j,k) \mid \exists (j',k') \in \mathcal{I}^+,\ j \le j', k \le k'\}$. Similarly define $S^- = \{(j,k) \mid \exists (j',k') \in \mathcal{I}^-,\ j \ge j', k \ge k'\}$. We only need to consider the case that the condition 3 of Lemma 2 is satisfied. Also, if S^+ or S^- is not monotone in the sense of conditions 1 and 2 of Lemma 2, we can reduce the L_1 distance between x and y. This can be seen as follows. If S^+ or S^- is not monotone, then we can find a pattern in Figure 5 (or a vertical pattern of this). Without loss of generality let $j_2 - j_1 \le j_4 - j_3$ and define $j_5 := j_4 - (j_2 - j_1)$. For simplicity assume $j_2 < j_3$ or $k_0 \ne k_3$. Then by applying

$$z_1 := -\boldsymbol{e}_{j_1k_0} + \boldsymbol{e}_{j_2k_0} - \boldsymbol{e}_{j_5k_1} + \boldsymbol{e}_{j_4k_1}$$

to z, we can reduce the L_1 distance between x and y by at least four. The case of $k_0 = k_1$

$$\begin{array}{c|cccc|} & j_1 & & & j_2 \\ \cline{2-5} k_0 & + & 0 & \cdots & 0 \quad - \\ \cline{2-5} \end{array} \qquad \begin{array}{c|cccc|} & j_3 & j_5 & & j_4 \\ \cline{2-5} k_1 & - \quad 0 & & \cdots & 0 \quad + \\ \cline{2-5} \end{array}$$

Figure 5: Case 1 or Case 2-1

and $j_2 = j_3$ needs a special consideration, but the monotonicity holds with respect to the horizontal separating line through (k_0, j_2). Therefore it suffices to consider the case that S^+ and S^- are monotone in the sense of conditions 1 and 2 of Lemma 2. Then

$$\sum_{j,k} (aj + bk + c) z_{1jk} = 0$$

implies that non-zero elements $z_{1jk} \neq 0$ only exist on the line $\{(j,k) \mid aj + bk + c\}$. Then the problem reduces to the univariate logistic regression.

Case 2. Next we consider the case that only one of the patterns of (14) or (15) exists. Without loss of generality, we assume that (15) holds and from Lemma 1 we assume that there exist $j_1 < j_2 \leq j_3 < j_4$ such that

$$z_{1j_1+} > 0, \quad z_{1j_2+} < 0, \quad z_{1j_3+} < 0, \quad z_{1j_4+} > 0.$$

In this case either a pattern of signs in Figure 5 or a pattern in Figure 6 has to exist in z^1.

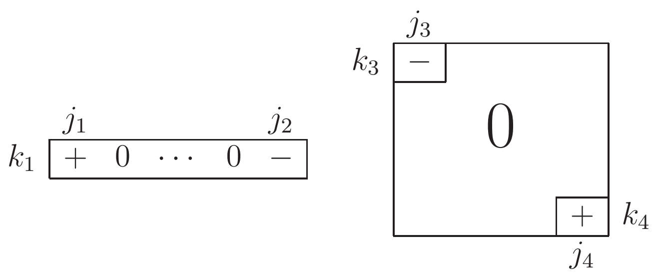


Figure 6: Case 2-2

Case 2-1. The case of Figure 5.
In this case we can reduce the L_1 distance between x and y as in Case 1.

Case 2-2. The case of Figure 6.
In the case of Figure 6, we distinguish two subcases depending on $j_2 - j_1 \leq j_4 - j_3$ or $j_2 - j_1 > j_4 - j_3$.

Case 2-2-1. $j_2 - j_1 \leq j_3 - j_4$.
Let $j_5 := j_3 + (j_2 - j_1)$. By applying

$$z_{2a} := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_2k_1} + \boldsymbol{e}_{j_3k_3} - \boldsymbol{e}_{j_5k_3}$$

to z, we reduce the L_1 distance by four.

Case 2-2-2. $j_2 - j_1 > j_4 - j_3$.
In this case we prove the theorem by induction on $j_4 - j_3$. When $j_4 - j_3 = 0$, the problem is reduced to Case 1. Therefore we assume that $j_4 - j_3 > 0$.

Case 2-2-2-1. $(x_{1,j_1+1,k_1}, x_{1,j_4-1,k_4}) > 0$ or $(x_{2,j_1+1,k_1}, x_{2,j_4-1,k_4}) > 0$.
In this case we can apply

$$z_{2b} := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1+1,k_1} + \boldsymbol{e}_{j_4-1,k_4} - \boldsymbol{e}_{j_4,k_4}$$

to z and then

$$y(j_3, j_4 - 1; k_3, k_4) \subset z + z_{2b}, \quad \|z + z_{2b}\|_1 = \|z\|_1.$$

Hence $\|z\|_1$ is can be reduced by moves of $\mathcal{B}_{\Lambda(A\otimes B)}$ from the inductive assumption.

In the case where $(x_{1,j_2-1,k_2}, x_{1,j_3+1,k_3}) > 0$ or $(x_{2,j_2-1,k_2}, x_{2,j_3+1,k_3}) > 0$, the proof is similar.

Case 2-2-2-2. $(x_{1,j_1+1,k_1}, x_{2,j_3+1,k_3}) > 0$ or $(x_{2,j_1+1,k_1}, x_{1,j_3+1,k_3}) > 0$,
In this case we can apply

$$z_{2c} := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1+1,k_1} + \boldsymbol{e}_{j_3k_3} - \boldsymbol{e}_{j_3+1,k_3}$$

to z and then

$$y(j_3 + 1, j_4; k_3, k_4) \subset z + z_{2c}.$$

Therefore $\|z\|_1$ can be reduced by moves of $\mathcal{B}_{\Lambda(A\otimes B)}$ from the inductive assumption.

In the case where $(x_{1j_2-1,k_2}, x_{2j_4-1,k_4}) > 0$ or $(x_{2j_2-1,k_2}, x_{1j_4-1,k_4}) > 0$, the proof is similar.

Case 2-2-2-3. $(x_{1,j_1+1,k_1}, x_{2,j_2-1,k_2}, x_{1,j_3+1,k_3}, x_{2,j_4-1,k_4}) = 0$.
In this case we have

$$(x_{2,j_1+1,k_1}, x_{1,j_2-1,k_2}, x_{2,j_3+1,k_3}, x_{1,j_4-1,k_4}) > 0.$$

Then there exists $j_1 < j_5 < j_2$ such that $(x_{2j_5k_1}, x_{1j_5+1,k_1}) > 0$. Hence we can apply

$$z_{2d}^1 := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1,k_1+1} + \boldsymbol{e}_{j_5k_1} - \boldsymbol{e}_{j_5+1,k_1}$$

and

$$z_{2d}^2 := -\boldsymbol{e}_{j_5k_1} + \boldsymbol{e}_{j_5+1,k_1} + \boldsymbol{e}_{j_4-1,k_4} - \boldsymbol{e}_{j_4k_4}$$

to z in this order and then we have

$$\|z + z_{2d}^1 + z_{2d}^2\|_1 = \|z\|_1 \quad \text{and} \quad y(j_1, k_1 + 1; j_2, k_2) \subset z + z_{2d}^1 + z_{2d}^2.$$

Hence theorem holds from the inductive assumption.

In the case where $(x_{2,j_1+1,k_1}, x_{1,j_2-1,k_2}, x_{2,j_3+1,k_3}, x_{1,j_4-1,k_4}) > 0$, the proof is similar.

Case 3. We now consider the case that there exist no j_0, k_1, k_2 satisfying (14) and there exist no k_0, j_1, j_2 satisfying (15). From Lemma 1, either of the patterns of signs as in Figure 7 and Figure 8 has to exist in z^1. Here we consider the case that patterns in Figure 7 exist. The case of Figure 8 will be treated as Case 5 below. We make various subcases depending on the sizes of two rectangles in Figure 7.

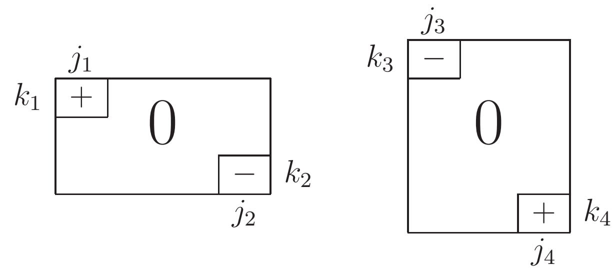


Figure 7: Case 3

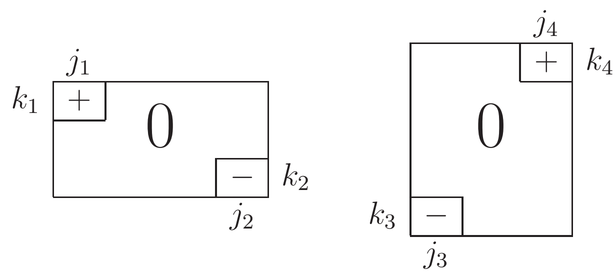


Figure 8: Case 5

Case 3-1. $j_2 - j_1 \geq j_4 - j_3$ and $k_2 - k_1 \geq k_4 - k_3$.
In this case the left rectangle contains the right rectangle in Figure 7. Define (j_5, k_5) by

$$(j_5, k_5) := (j_1, k_1) - (j_3, k_3) + (j_4, k_4).$$

Then

$$z_{3a} := -\boldsymbol{e}_{j_1 k_1} + \boldsymbol{e}_{j_5 k_5} + \boldsymbol{e}_{j_3 k_3} - \boldsymbol{e}_{j_4 k_4}$$

reduces the L_1 distance by four.

In the case where $j_2 - j_1 \leq j_4 - j_3$ and $k_2 - k_1 \leq k_4 - k_3$, the proof is similar.

Case 3-2. When there is no inclusion relation between two rectangles of Figure 7, it suffices to consider the case of $j_2 - j_1 > j_4 - j_3$ and $k_2 - k_1 < k_4 - k_3$. We prove the theorem by induction on $l := (k_2 - k_1) + (j_4 - j_3)$. If $l = 0$, the theorem holds from Case 1.

Case 3-2-1. $(x_{1 j_1, k_1+1}, x_{1 j_4, k_4-1}) > 0$.
In this case we can apply

$$z_{3b} := -\boldsymbol{e}_{j_1 k_1} + \boldsymbol{e}_{j_1, k_1+1} + \boldsymbol{e}_{j_4, k_4-1} + \boldsymbol{e}_{j_4 k_4}$$

and then we have

$$y(j_1, j_2; k_1 + 1, k_2) \subset z + z_{3b}, \quad y(j_3, j_4; k_4, k_4 - 1) \subset z + z_{3b}.$$

From the inductive assumption the theorem holds in this case.

Also in the following cases, the proof is similar.

- $(x_{2j_1,k_1+1}, x_{2j_4,k_4-1}) > 0$;
- $(x_{i,j_1+1,k_1}, x_{i,j_4-1,k_4}) > 0$;
- $(x_{ij_2,k_2-1}, x_{ij_3,k_3+1}) > 0$;
- $(x_{i,j_2-1,k_2}, x_{i,j_3+1,k_3}) > 0$;

Case 3-2-2. $(x_{1j_1,k_1+1}, x_{2j_3,k_3+1}) > 0$.

In this case we can apply

$$z_{3c} := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1,k_1+1} + \boldsymbol{e}_{j_3k_3} + \boldsymbol{e}_{j_3,k_3+1}$$

and then we have

$$y(j_1, j_2; k_1+1, k_2) \subset z + z_{3c}, \quad y(j_3, j_4; k_3+1, k_4) \subset z + z_{3c}.$$

From the inductive assumption the theorem holds in this case.

Also in the following cases, the proof is in the similar way.

- $(x_{2,j_1,k_1+1}, x_{1,j_3,k_3+1}) > 0$;
- $(x_{i,j_1+1,k_1}, x_{i^*,j_3+1,k_3}) > 0$;
- $(x_{i,j_2,k_2-1}, x_{i^*,j_4,k_4-1}) > 0$;
- $(x_{i,j_2-1,k_2}, x_{i^*,j_4-1,k_4}) > 0$,

where $i^* := 3 - i$.

Case 3-2-3. $x_{2j_1,k_1+1} = x_{1j_2,k_2-1} = x_{2j_3,k_3+1} = x_{1j_4,k_4-1} = 0$.

From the result of Case 3-2-1 and Case 3-2-2, it suffices to consider the case where

$$x_{2j_1,k_1+1} = x_{1j_2,k_2-1} = x_{2j_3,k_3+1} = x_{1j_4,k_4-1} = 0. \tag{16}$$

We note that (16) implies

$$(x_{1j_1,k_1+1}, x_{2j_2,k_2-1}, x_{1j_3,k_3+1}, x_{2j_4,k_4-1}) > 0.$$

Therefore there exist $j_1 < j_5 < j_2$ and $k_1 < k_5 < k_2$ satisfying either

$$x_{1j_5k_5} > 0 \quad x_{2j_5,k_5+1} > 0 \tag{17}$$

or

$$x_{1j_5k_5} > 0 \quad x_{2,j_5+1,k_5} > 0. \tag{18}$$

Case 3-2-3-1. $x_{1j_5k_5} > 0$ and $x_{2j_5,k_5+1} > 0$ (17).

In this case we can apply

$$z_{3d}^1 := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1,k_1+1} + \boldsymbol{e}_{j_5k_5} - \boldsymbol{e}_{j_5,k_5+1}$$

and

$$z_{3d}^2 := -\boldsymbol{e}_{j_5k_5} + \boldsymbol{e}_{j_5,k_5+1} + \boldsymbol{e}_{j_4k_4-1} - \boldsymbol{e}_{j_4k_4}$$

to z in this order. Then we have $\|z + z_{3d}^1 + z_{3d}^2\|_1 = \|z\|_1$ and

$$y(j_1, j_2; k_1+1, k_2) \subset z + z_{3d}^1 + z_{3d}^2, \quad y(j_3, j_4; k_3, k_4-1) \subset z + z_{3d}^1 + z_{3d}^2.$$

Hence from the inductive assumption the L_1 distance can be reduced by moves in $\mathcal{B}_{\Lambda(A\otimes B)}$.

Case 3-2-3-2. $x_{1j_5k_5} > 0$ and $x_{2,j_5+1,k_5} > 0$ (18).

In this case we further consider subcases depending on the value of $x_{1,j_1+1,k}$.

Case 3-2-3-2-1. $x_{1,j_1+1,k_1} > 0$.

From the result of Case 3-2-1 and and Case 3-2-2, it suffices to consider the case where

$$x_{2,j_1+1,k_1} = x_{1,j_4-1,k_4} = x_{2,j_3+1,k_3} = x_{1,j_2-1,k_2} = 0. \tag{19}$$

We note that (19) implies that

$$(x_{1,j_1+1,k_1}, x_{2,j_4-1,k_4}, x_{1,j_3+1,k_3}, x_{2,j_2-1,k_2}) > 0.$$

Since (18) is satisfied, we can apply

$$z_{3e}^1 := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1+1,k_1} + \boldsymbol{e}_{j_5k_5} - \boldsymbol{e}_{j_5+1,k_5}$$

and

$$z_{3e}^2 := -\boldsymbol{e}_{j_5k_5} + \boldsymbol{e}_{j_5+1,k_5} + \boldsymbol{e}_{j_4-1,k_4} - \boldsymbol{e}_{j_4k_4}$$

in this order. Then we have $\|z + z_{3e}^1 + z_{3e}^2\|_1 = \|z\|_1$ and

$$y(j_1+1, j_2; k_1, k_2) \subset z + z_{3e}^1 + z_{3e}^2, \quad y(j_3, j_4-1; k_3, k_4) \subset z + z_{3e}^1 + z_{3e}^2.$$

Hence from the inductive assumption, L_1 distance can be reduced by moves in $\mathcal{B}_{\Lambda(A\otimes B)}$.

If any of $x_{2,j_2-1,k_2}, x_{1,j_3+1,k_3}, x_{2,j_4-1,k_4}$ is positive, the same argument can be applied.

Case 3-2-3-2-2. (Case 4) $x_{1,j_1+1,k_1} = x_{2,j_2-1,k_2} = x_{1,j_3+1,k_3} = x_{2,j_4-1,k_4} = 0$.

For readability, we relabel this case as Case 4. In this case

$$(x_{2,j_1+1,k_1}, x_{1,j_2-1,k_2}, x_{2,j_3+1,k_3}, x_{1,j_4-1,k_4}) > 0.$$

Then there exists $j_1 < j_6 < j_2$ and $k_1 < k_6 < k_2$ satisfying either

$$x_{2j_6k_6} > 0, \quad x_{1,j_6+1,k_6} > 0 \tag{20}$$

or

$$x_{2j_6k_6} > 0, \quad x_{1,j_6,k_6+1} > 0. \tag{21}$$

Case 4-1. The case that (20) is satisfied.

In this case the proof is in similar to Case 3-2-3-1.

Case 4-2. The case that (20) is not satisfied.

In this case

$$x_{1jk_1} = 0, \quad j = j_1 + 1, \ldots, j_2, \quad x_{2j_3k} = 0, \quad k = k_3 + 1, \ldots, k_4. \tag{22}$$

We can assume without loss of generality that $j_2 < j_3$. Then we note that

$$(j_7, k_7) := (j_4, k_4) - (j_2, k_2) + (j_1, k_1) \in \mathcal{J}$$

where $\mathcal{J} := [J] \times [K]$.

Case 4-2-1. $x_{2j_7k_7} > 0$ or $y_{1j_7k_7} > 0$.

In this case we can apply

$$z_{4a} := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_2k_2} + \boldsymbol{e}_{j_7k_7} - \boldsymbol{e}_{j_4k_4}$$

to z and we can reduce the L_1 distance by four.

Case 4-2-2. $x_{2j_7k_7} = 0$ and $y_{1j_7k_7} = 0$.

In this case we have $z_{1j_7k_7} > 0$.

Case 4-2-2-1. The case that there exists $j_7 < j_8 < j_3$ such that $z_{1j_8k_7} < 0$.

In this case we can prove the theorem in the same way as Case 2-2.

Case 4-2-2-2. The case that $z_{1jk_7} \geq 0$ for all $j_7 < j < j_3$.

From the condition (22) there exists j_9 satisfying either of the following conditions,

(i) $j_7 \leq j_9 < j_3$, $z_{1j_9k_7} > 0$ and $x_{1,j_9+1,k_7} > 0$;

(ii) $j_7 < j_9 < j_3$, $z_{1j_9k_7} = z_{1j_9+1,k_7} = 0$, $x_{2j_9k_7} > 0$ and $x_{1j_9+1,k_7} > 0$.

Case 4-2-2-2-1. The case that (i) is satisfied.

In this case by applying the move

$$z_{4b} := -\boldsymbol{e}_{j_9k_7} + \boldsymbol{e}_{j_9+1,k_7} + \boldsymbol{e}_{j_4-1,k_4} - \boldsymbol{e}_{j_4k_4},$$

we have $\|z + z_{4b}\|_1 = \|z\|_1$ and

$$y(j_1, j_2; k_1, k_2) \subset z + z_{4b}, \quad y(j_3, j_4 - 1; k_3, k_4) \subset z + z_{4b}.$$

Hence the theorem holds from the inductive assumption.

Case 4-2-2-2-2. The case that (ii) is satisfied.

In this case by applying the move z_{4b} and

$$z_{4c} := -\boldsymbol{e}_{j_1k_1} + \boldsymbol{e}_{j_1+1,k_1} + \boldsymbol{e}_{j_9k_7} + \boldsymbol{e}_{j_9+1,k_7}$$

in this order and then we have $\|z + z_{4b} + z_{4c}\|_1 = \|z\|_1$ and

$$y(j_1+1, j_2; k_1, k_2) \subset z + z_{4b} + z_{4c}, \quad y(j_3, j_4-1; k_3, k_4) \subset z + z_{4b} + z_{4c}$$

Hence the theorem holds from the inductive assumption.

Case 5. We now consider the case where z^1 contains patterns of signs in Figure 8 and does not contain patterns of signs in Figure 7. We show that if z_1 contains the pattern of signs in Figure 9, we can reduce the L_1 norm z or otherwise z is not a move. The proof is by induction on

$$l := \min\left((j_2 - j_1) + (k_2 - k_1), (j_4 - j_3) + (k_3 - k_4)\right).$$

When $l = 1$, theorem holds by Case 1.

Case 5-1. $k_2 > k_4$

Based on the argument in Case 3-2-1 and 3-2-2, we only need to consider the case that z^1 contains patterns in Figure 9, where $z_{1jk} = 0^*$ and $z_{1jk} = 0_*$ denote $x_{1jk} = y_{1jk} > 0$ and $x_{2jk} = y_{2jk} > 0$, respectively. Define two set of cells $\mathcal{A}_1$ and $\mathcal{A}_2$ as in Figure 9. Then there exist j_5, k_5, j_6 and k_6 such that

$$j_1 < j_5 < j_2, \quad j_3 < j_6 < j_4, \quad k_1 < k_5 < k_2, \quad k_4 < k_6 < k_4,$$

$$z_{1j_5k_5} = 0^*, \quad z_{1j_5k_5+1} = 0_*, \quad z_{1j_6k_6} = 0_*, \quad z_{1j_6,k_6+1} = 0^*$$

as represented in Figure 9(i). Then we can apply the move

$$z_{5a} := \boldsymbol{e}_{j_5k_5} - \boldsymbol{e}_{j_5,k_5+1} - \boldsymbol{e}_{j_6k_6} - \boldsymbol{e}_{j_5,k_6+1},$$

$$z_{5b} := \boldsymbol{e}_{j_6k_6} - \boldsymbol{e}_{j_6,k_6+1} - \boldsymbol{e}_{j_4k_4} + \boldsymbol{e}_{j_4,k_4+1}$$

to z in this order and $z' := z + z_{5a} + z_{5b}$ is expressed as in Figure 9(ii). Suppose that there exists $(j,k) \in \mathcal{A}_1$ such that $z_{1jk} < 0$. Then $z_{1jk_4} = 0$ and hence there exists $k \le k' < k_4$ such that

$$z_{1jk'} < 0, \quad z_{1j,k'+1} = 0.$$

Therefore we can apply the move

$$z_{5c} := -\boldsymbol{e}_{j_5k_5} + \boldsymbol{e}_{j_5,k_5+1} + \boldsymbol{e}_{jk'} - \boldsymbol{e}_{j,k'+1}$$

and $z'' := z' + z_{5c}$ satisfies

$$\|z''\|_1 = \|z\|_1, \quad y(j_1, j_2; k_1, k_2) \subset z''.$$

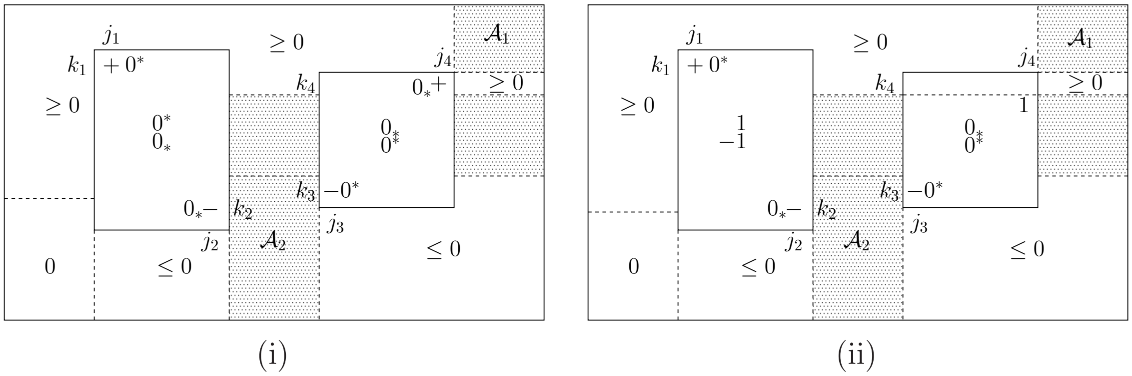


Figure 9: Case 5-1

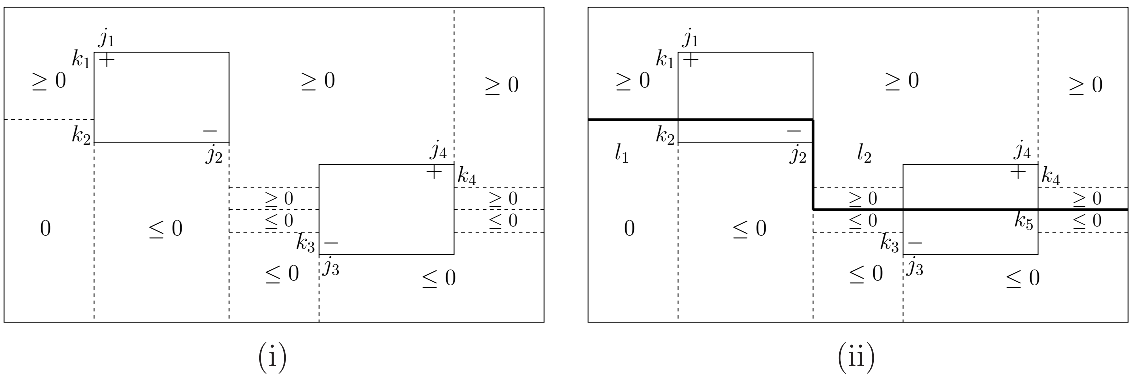


Figure 10: Case 5-2

Therefore the theorem holds by the inductive assumption.

Similarly we can prove the theorem in the case where there exists $(j,k) \in \mathcal{A}_2$ such that $z_{1jk} > 0$.

Now we suppose that $z_{1jk} \geq 0$ for $(j,k) \in \mathcal{A}_1$ and $z_{1jk} \leq 0$ for $(j,k) \in \mathcal{A}_1$. Since there does not exist the pattern in Figure 7, there exist $k_4 < k_7 < k_3$ such that $z_{1jk} \geq 0$ for $k \leq k_7$ and $z_{1jk} \leq 0$ for $k > k_7$. This contradicts the condition $\sum_{k=1}^{K} k z_{1jk} = 0$ and hence z is not a move.

Case 5-2. $k_2 < k_4$

By using the same argument, we only need to consider the case that z^1 contains patterns in Figure 10(i). Then both S^+ and S^- is monotone in the sense of conditions 1 and 2 of Lemma 2. Therefore if z is a move, we can reduce L_1 norm of z from Lemma 2 in the similar way to Case 5-1.

Acknowledgments. We are grateful to Kazuo Murota for valuable comments on the separation lemma and to Ian Dinwoodie for constructive suggestion on the numerical examples.

References

Alan Agresti. *Categorical Data Analysis*. New York : John Wiley and Sons, 1st edition, 1990.

Satoshi Aoki and Akimichi Takemura. Markov chain Monte Carlo exact tests for incomplete two-way contingency table. *Journal of Statistical Computation and Simulation*, 75(10):787–812, 2005.

Satoshi Aoki, Takayuki Hibi, Hidefumi Ohsugi, and Akimichi Takemura. Gröbner bases of nested configurations. *Journal of Algebra*, 320:2583–2593, 2008.

N.E. Breslow and N.E. Day. *Statistical Methods in Cancer Research (Vol. 1): The Analysis of Case-control Studies*. Lyon : International Agency for Research on Cancer, 1980.

Y. Chen, I. Dinwoodie, and R. Yoshida. Markov chains, quotient ideals, and connectivity with positive margins. Algebraic and Geometric Methods in Statistics *dedicated to Professor Giovanni Pistone (P. Gibilisco, E. Riccomagno, M.-P. Rogantin, H. P. Wynn, eds.)*, 2008. To appear.

Yuguo Chen, Ian Dinwoodie, Adrian Dobra, and Mark Huber. Lattice points, contingency tables, and sampling. In *Integer points in polyhedra—geometry, number theory, algebra, optimization*, volume 374 of *Contemp. Math.*, pages 65–78. Amer. Math. Soc., Providence, RI, 2005.

Ronald Christensen. *Log-Linear Models and Logistic Regression*. New York : Springer Verlag, 2nd edition, 1997.

J. Cornfield. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. *Fed. Proc.*, 21(11):58–61, 1962.

Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1):363–397, 1998. ISSN 0090-5364.

Persi Diaconis, David Eisenbud, and Bernd Sturmfels. Lattice walks and primary decomposition. In *Mathematical essays in honor of Gian-Carlo Rota (Cambridge, MA, 1996)*, volume 161 of *Progr. Math.*, pages 173–193. Birkhäuser Boston, Boston, MA, 1998.

Sigeiti Moriguti. A modification of Schwarz's inequality with applications to distributions. *The Annals of Mathematical Statistics*, 24(1):107–113, 1953.

Bernd Sturmfels. *Gröbner Bases and Convex Polytopes*, volume 8 of *University Lecture Series*. American Mathematical Society, Providence, RI, 1996. ISBN 0-8218-0487-1.

Akimichi Takemura and Satoshi Aoki. Distance reducing Markov bases for sampling from a discrete sample space. *Bernoulli*, 11(5):793–813, 2005.