

Random Generation of $2 \times \dots \times 2 \times J$ Contingency Tables

Tomomi MATSUI¹, Yasuko MATSUI², Yoko ONO³

¹Department of Mathematical Informatics,
Graduate School of Information Science and Technology,
University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan.
<http://www.misojiro.t.u-tokyo.ac.jp/~tomomi/>

²Department of Mathematical Sciences, School of Science,
Tokai University, Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan.
yasuko@ss.u-tokai.ac.jp

³Department of Management Science, Faculty of Engineering,
Tokyo University of Science, Shinjuku-ku, Tokyo 162-8601, Japan.
ono@ms.kagu.tus.ac.jp

METR 2003-03 (January 2003)

Abstract

We propose two Markov chains for sampling $(m + 1)$ -dimensional contingency tables indexed by $\{1, 2\}^m \times \{1, 2, \dots, n\}$. Stationary distributions of our chains are the uniform distribution and a conditional multinomial distribution (which is equivalent to the hypergeometric distribution if $m = 1$), respectively. The mixing times of our chains are bounded by $(1/2)n(n - 1) \ln(dn/\varepsilon)$ where d is the average of the values in cells and ε is a given error bound. We use the path coupling method for estimating the mixing time of our chains and showed that our chains are rapidly mixing.

1 Introduction

We propose two Markov chains for sampling $(m+1)$ -dimensional contingency tables indexed by $\{1, 2\}^m \times \{1, 2, \dots, n\}$. First chain has the uniform distribution as a unique stationary distribution. The stationary distributions of second chain is a conditional multinomial distribution, which is the hypergeometric distribution for 2-dimensional case. The mixing times of our chains are bounded by $(1/2)n(n-1)\ln(dn/\varepsilon)$ where d is the average of the values in cells and ε is a given error bound. We use the path coupling method [5, 6] for estimating the mixing time of our chains.

Contingency tables are used in statistics to store data from sample surveys. Consider a scenario where N subjects are categorized into a table according to some attributes. Data is often analyzed under assumption that the attributes are independent; that is, the joint distribution is uniquely determined by the marginal probabilities. We often assume that each table was generated from the uniform distribution, or a conditional multinomial distribution (hypergeometric distribution) over the set of all the contingency tables (see [1, 2, 8, 15] for example). One of the commonly used measure of independence is the χ^2 statistics [23]. A typical test of the independence asks what fraction (the sum of probabilities) of tables have χ^2 value smaller than a parameter t , as t varies. When the marginal totals are sufficiently large, we can apply the Pearson chi-square test [23]. In case that marginal totals includes a small number, we need an exact inference for contingency tables [15]. For the analysis of 2×2 contingency tables, an alternative to maximum likelihood estimation and χ^2 goodness-of fit tests is the use of Fisher's exact test for independence [16].

Exact test can be done by systematic enumeration of all the tables. When the number of tables is huge, exact enumeration is impractical. Mehta and Patel [22] proposed a network algorithm for exact counting (not for enumeration) of contingency tables. However, the computational efforts and memory requirement of their algorithm is bounded by the table sum and so impractical when the table sum is large. For estimating the moments of χ^2 statistics efficiently, a standard technique is the ordinary Monte Carlo method if we have a method for sampling from the set of contingency tables. By using a rapidly mixing Markov chain with the desired stationary distribution, we can sample a contingency table after enough number of transitions of the Markov chain from arbitrary initial state.

It is known that the problem for generating 3-dimensional contingency tables is intractable. More precisely, when we deal with 3-dimensional tables, the problem for checking the existence of at least one table satisfying the given marginal totals is NP-complete [18]. Diaconis and Strumfels [11] proposed an algorithm for finding a Markov base for higher dimensional contingency tables. Recently, Aoki and Takemura discussed Markov bases for some classes of 3-dimensional contingency tables [4, 25]. In this paper, we deal with a special class of $(m+1)$ -dimensional contingency tables such that the cells are indexed by $\{1, 2\}^m \times \{1, 2, \dots, n\}$. For this class, a natural Markov basis exists, which is a direct extension of 2-dimensional case. This class of contingency tables arises in many practical situations [14, 24]. There also exist some theoretical results on testing the independency of attributes of $2 \times 2 \times K$ tables (see Agresti's survey paper [1] for example).

The problem of almost uniform sampling of contingency tables can be solved by using a Markov chain that converges to the uniform distribution. Diaconis and Saloff-Coste [10] discussed the rate of convergence of a natural Markov chain for 2-dimensional contingency tables. They have shown that the ordinary chain mixes polynomial time in the table sum when the numbers of rows and columns are

fixed. Dyer, Kannan and Mount [13] proposed a different Markov chain for counting the number of 2-dimensional contingency tables. In case of sufficiently large marginal totals, their chain mixes polynomial time in the number of rows and columns. For 2-dimensional contingency tables with two rows, Hernek [17] showed that the mixing time of the ordinary Markov chain is bounded by a polynomial of table sum and number of columns. Hernek bounded the mixing time of the chain by using coupling lemma shown by Aldous [3]. Dyer and Greenhill [12] proposed a rapidly mixing Markov chain for two rowed contingency tables. Their chain mixes polynomial time in the logarithm of table sum and the number of columns. They analyzed the mixing rate of their chain by using path coupling technique proposed by Bublely and Dyer [5, 6]. Kannan, Tetali and Vempala [21] gave a Markov chain with polynomial-time convergence for the 0-1 case with nearly equal marginal totals. In contrast, Chung, Graham and Yau [7] proposed a Markov chain for contingency tables with large enough marginal sums and showed that their chain converges in pseudo polynomial time.

We also consider the problem for generating contingency tables from a conditional multinomial distribution (hypergeometric distribution) over the set of all the contingency tables. In 2-dimensional case, there exists a simple $O(N)$ time perfect sampling method, where N is the table sum. However, 3-dimensional case is computationally intractable in general, as described above. It is easy to see that the problem for sampling perfect matchings in a given bipartite graph uniformly is a special case of the problem for generating $2 \times I \times J$ contingency tables from a conditional multinomial distribution. The existence of a polynomial time approximate uniform sampler for perfect matching was a long standing open problem (see [19] for example). Recently, the problem was solved affirmatively [20].

In Section 2, we introduce some notations and summarize the path coupling method. In Section 3, we describe our first chain whose stationary distribution is uniform. Section 4, we discuss our second chain whose stationary distribution is a conditional multinomial distribution.

2 Notations and Definitions

We denote the set of integers (non-negative integers, positive integers) by \mathbb{Z} (\mathbb{Z}_+ , \mathbb{Z}_{++}), respectively. In this paper, we consider a set of $(m + 1)$ -dimensional contingency tables indexed by $\mathbb{B}^m \times J$ where $\mathbb{B} = \{1, 2\}$ and $J = \{1, 2, \dots, n\}$. Any index in J is called a *column index*. For any vector $\mathbf{x} \in \mathbb{Z}^{\mathbb{B}^m \times J}$, both $x(\mathbf{i}; j)$ and $x(i_1, i_2, \dots, i_m; j)$ denote the element of \mathbf{x} indexed by $\mathbf{i} = (i_1, i_2, \dots, i_m) \in \mathbb{B}^m$ and $j \in J$. For any column index $j \in J$, $\mathbf{x}(j) \in \mathbb{Z}^{\mathbb{B}^m}$ denotes the subvector of $\mathbf{x} \in \mathbb{Z}^{\mathbb{B}^m \times J}$ consists of elements defined by indices in $\mathbb{B}^m \times \{j\}$. Given a vector of indices $\mathbf{i} \in \mathbb{B}^m$ and an index $l \in \{1, 2, \dots, m\}$, $\mathbf{i}_{\overline{l}}$ denotes the vector of indices $(i_1, \dots, i_{l-1}, i_{l+1}, \dots, i_m) \in \mathbb{B}^{m-1}$ and we also denote the vector \mathbf{i} by $(\mathbf{i}_{\overline{l}}, i_l)$ by changing the order of elements. For any vector $\mathbf{x} \in \mathbb{Z}^{\mathbb{B}^m \times J}$ and $l \in \{1, 2, \dots, m\}$, $x(\mathbf{i}_{\overline{l}}, i_l; j)$ denotes the element $x(\mathbf{i}; j)$.

Let $(\mathbf{r}^1, \mathbf{r}^2, \dots, \mathbf{r}^m; \mathbf{c})$ be a sequence of non-negative integer vectors where $\mathbf{r}^l \in \mathbb{Z}_+^{\mathbb{B}^{m-1} \times J}$ for each $l \in \{1, 2, \dots, m\}$ and $\mathbf{c} \in \mathbb{Z}_+^{\mathbb{B}^m}$. The element of \mathbf{r}^l indexed by $(\mathbf{i}'; j) \in \mathbb{B}^{m-1} \times J$ is denoted by $r^l(\mathbf{i}'; j)$. The set of contingency tables corresponding to $(\mathbf{r}^1, \mathbf{r}^2, \dots, \mathbf{r}^m; \mathbf{c})$ is defined by

$$\mathcal{T} \stackrel{\text{def.}}{=} \left\{ \mathbf{x} \in \mathbb{Z}_+^{\mathbb{B}^m \times J} \left| \begin{array}{l} x(\mathbf{i}_{\overline{l}}, 1; j) + x(\mathbf{i}_{\overline{l}}, 2; j) = r^l(\mathbf{i}_{\overline{l}}, j) \quad (\forall l \in \{1, 2, \dots, m\}, \forall \mathbf{i}_{\overline{l}} \in \mathbb{B}^{m-1}, \forall j \in J), \\ \sum_{j \in J} x(\mathbf{i}; j) = c(\mathbf{i}) \quad (\forall \mathbf{i} \in \mathbb{B}^m) \end{array} \right. \right\}.$$

Each element in \mathcal{T} is called a *table* for simplicity. In the following, $\sum_{\mathbf{i} \in \mathbb{B}^m} c(\mathbf{i})$ is denoted by N . Clearly,

$x(1, 1; 1)$	$x(1, 1; 2)$	$x(1, 1; 3)$	$x(1, 1; 4)$	$x(1, 1; 5)$	$x(1, 1; 6)$
$x(1, 2; 1)$	$x(1, 2; 2)$	$x(1, 2; 3)$	$x(1, 2; 4)$	$x(1, 2; 5)$	$x(1, 2; 6)$

2	5	1	4	5	9
6	3	6	3	2	0

$x(2, 1; 1)$	$x(2, 1; 2)$	$x(2, 1; 3)$	$x(2, 1; 4)$	$x(2, 1; 5)$	$x(2, 1; 6)$
$x(2, 2; 1)$	$x(2, 2; 2)$	$x(2, 2; 3)$	$x(2, 2; 4)$	$x(2, 2; 5)$	$x(2, 2; 6)$

5	2	7	1	0	3
6	4	4	3	8	8

 $c(1, 1)$
 $c(1, 2)$

8	8	7	7	7	9
---	---	---	---	---	---

 $c(2, 1)$
 $c(2, 2)$

11	6	11	4	8	11
----	---	----	---	---	----

$r^2(1; 1)$	$r^2(1; 2)$	$r^2(1; 3)$	$r^2(1; 4)$	$r^2(1; 5)$	$r^2(1; 6)$
-------------	-------------	-------------	-------------	-------------	-------------

$r^1(1; 1)$	$r^1(1; 2)$	$r^1(1; 3)$	$r^1(1; 4)$	$r^1(1; 5)$	$r^1(1; 6)$
$r^1(2; 1)$	$r^1(2; 2)$	$r^1(2; 3)$	$r^1(2; 4)$	$r^1(2; 5)$	$r^1(2; 6)$

7	7	8	5	5	12
12	7	10	6	10	8

Figure 1: An example of $B \times B \times J$ ($|J| = 6$) table (denoted by \mathbf{x}^*).

for any table $\mathbf{x} \in \mathcal{T}$, the sum total of elements of \mathbf{x} is equal to N .

In the rest of this section, we briefly review the path coupling technique proposed by Bubley and Dyer [5]. We use the technique in later sections to estimate the mixing time of our Markov chains. Here we deal with a Markov chain \mathcal{M} with state space \mathcal{T} . Assume that \mathcal{M} has a unique stationary distribution $\pi : \mathcal{T} \rightarrow [0, 1]$. For any probability distribution function π' on \mathcal{T} , define the *total variation distance* between π and π' to be

$$D_{\text{TV}}(\pi, \pi') \stackrel{\text{def.}}{=} \max_{\mathcal{T}' \subseteq \mathcal{T}} \left| \sum_{\mathbf{x} \in \mathcal{T}'} \pi(\mathbf{x}) - \sum_{\mathbf{x} \in \mathcal{T}'} \pi'(\mathbf{x}) \right| = (1/2) \sum_{\mathbf{x} \in \mathcal{T}} |\pi(\mathbf{x}) - \pi'(\mathbf{x})|.$$

If the initial state of the chain \mathcal{M} is \mathbf{x} , we denote the distribution of the chain at time t by $P_{\mathbf{x}}^t : \mathcal{T} \rightarrow [0, 1]$, i.e.,

$$P_{\mathbf{x}}^t(\mathbf{y}) \stackrel{\text{def.}}{=} \Pr[X_t = \mathbf{y} \mid X_0 = \mathbf{x}] \quad (\forall \mathbf{y} \in \mathcal{T}).$$

The rate of convergence to stationary from the initial state \mathbf{x} may be measured by

$$\tau_{\mathbf{x}}(\varepsilon) \stackrel{\text{def.}}{=} \min\{t \mid D_{\text{TV}}(\pi, P_{\mathbf{x}}^t) \leq \varepsilon \text{ for all } t' \geq t\}$$

where the error bound ε is a given positive constant. The *mixing time* $\tau(\varepsilon)$ of \mathcal{M} is defined by

$$\tau(\varepsilon) \stackrel{\text{def.}}{=} \max_{\mathbf{x} \in \mathcal{T}} \tau_{\mathbf{x}}(\varepsilon),$$

which is independent of the initial state.

Next, we define a special Markov process with respect to \mathcal{M} called joint process. A *joint process* of \mathcal{M} is a Markov chain (X_t, Y_t) defined on $\mathcal{T} \times \mathcal{T}$ satisfying that each of $(X_t), (Y_t)$, considered marginally, is a faithful copy of the original Markov chain \mathcal{M} . More precisely, we require that

$$\begin{aligned} \Pr[X_{t+1} = \mathbf{x}' \mid (X_t, Y_t) = (\mathbf{x}, \mathbf{y})] &= P_{\mathcal{M}}(\mathbf{x}, \mathbf{x}'), \\ \Pr[Y_{t+1} = \mathbf{y}' \mid (X_t, Y_t) = (\mathbf{x}, \mathbf{y})] &= P_{\mathcal{M}}(\mathbf{y}, \mathbf{y}'), \end{aligned}$$

for all $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathcal{T}$ where $P_{\mathcal{M}}(\mathbf{x}, \mathbf{x}')$ and $P_{\mathcal{M}}(\mathbf{y}, \mathbf{y}')$ denotes the transition probability from \mathbf{x} to \mathbf{x}' and from \mathbf{y} to \mathbf{y}' of the original Markov chain \mathcal{M} , respectively.

Path coupling lemma [Bubley and Dyer [5]]

Let G be a directed graph with vertex set \mathcal{T} and arc set $A \subseteq \mathcal{T} \times \mathcal{T}$. Let $\ell : A \rightarrow \mathbb{Z}_{++}$ be a positive length defined on the arc set. We assume that G is strongly connected. For any ordered pair of vertices $(\mathbf{x}, \mathbf{x}')$ of G , the distance from \mathbf{x} to \mathbf{x}' , denoted by $\ell(\mathbf{x}, \mathbf{x}')$, is the length of the shortest path from \mathbf{x} to \mathbf{x}' , where the length of a path is the sum of the lengths of arcs in the path. Suppose that there exists a joint process $(X, Y) \mapsto (X', Y')$ with respect to \mathcal{M} satisfying that

$$1 > \exists \beta > 0, \forall (X, Y) \in A, \mathbb{E}[\ell(X', Y')] \leq \beta \ell(X, Y).$$

Then the mixing time $\tau(\varepsilon)$ of the original Markov chain \mathcal{M} satisfies $\tau(\varepsilon) \leq (1 - \beta)^{-1} \ln(D/\varepsilon)$ where D denotes the diameter of G , i.e., the distance of a farthest (ordered) pair of vertices.

3 Markov Chain for Uniform Distribution

First, we show a lemma which implies an irreducible Markov chain defined on the set of tables \mathcal{T} . We define the *parity function* $p : \mathbb{Z} \rightarrow \{1, -1\}$ by

$$p(x) = \begin{cases} 1 & (x \text{ is an even integer}), \\ -1 & (x \text{ is an odd integer}). \end{cases}$$

For any index $\mathbf{i} \in \mathbb{B}^m$, we denote $p(i_1 + i_2 + \dots + i_m)$ by $p(\mathbf{i})$. The vector $\Delta \in \{1, -1\}^{\mathbb{B}^m}$ is defined by $\Delta(\mathbf{i}) \stackrel{\text{def.}}{=} p(\mathbf{i})$ for each index $\mathbf{i} \in \mathbb{B}^m$. Given an ordered pair of distinct column indices (j', j'') , we define the vector $\Delta[j', j''] \in \mathbb{Z}^{\mathbb{B}^m \times J}$ by

$$\Delta[j', j''](j) \stackrel{\text{def.}}{=} \begin{cases} \mathbf{0} & (j \in J \setminus \{j', j''\}), \\ \Delta & (j = j'), \\ -\Delta & (j = j''). \end{cases}$$

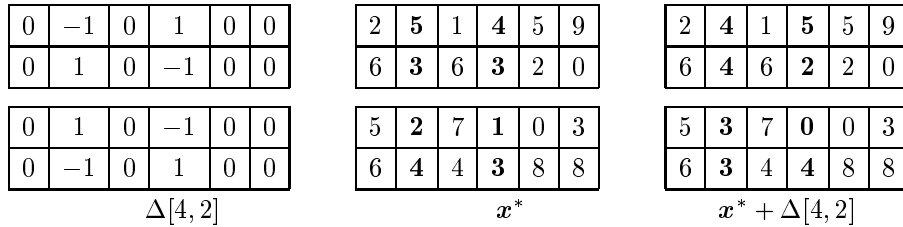


Figure 2: The vector $\Delta[4, 2]$.

For any table $\mathbf{x} \in \mathcal{T}$, we introduce the set of neighboring tables;

$$\mathbb{N}^0(\mathbf{x}) \stackrel{\text{def.}}{=} \{\mathbf{x}' \in \mathcal{T} \mid \exists (j', j'') \in J \times J, j' \neq j'', \mathbf{x}' = \mathbf{x} + \Delta[j', j'']\}.$$

It is easy to see that if $\mathbf{x}' = \mathbf{x} + \Delta[j', j'']$, then $\mathbf{x} = \mathbf{x}' - \Delta[j', j''] = \mathbf{x}' + \Delta[j'', j']$, and so $\mathbf{x}' \in \mathbb{N}^0(\mathbf{x})$ implies $\mathbf{x} \in \mathbb{N}^0(\mathbf{x}')$. For any pair of vectors $\mathbf{x}, \mathbf{x}' \in \mathbb{Z}^{\mathbb{B}^m \times J}$, $\|\mathbf{x} - \mathbf{x}'\|_1$ denotes the distance

$$\sum_{(\mathbf{i}; j) \in \mathbb{B}^m \times J} |\mathbf{x}(\mathbf{i}; j) - \mathbf{x}'(\mathbf{i}; j)| \text{ between } \mathbf{x} \text{ and } \mathbf{x}'.$$

Lemma 1 Let G^0 be an undirected graph with vertex set \mathcal{T} and for any pair of vertices $\{\mathbf{x}, \mathbf{x}'\}$, there exists an edge between \mathbf{x} and \mathbf{x}' if and only if $\mathbf{x}' \in \mathbb{N}^0(\mathbf{x})$. Then the graph G^0 is connected, i.e., for any pair of vertices $\{\mathbf{x}, \mathbf{x}'\}$ of G^0 , there exists a path on G^0 between \mathbf{x} and \mathbf{x}' . The diameter (the distance of farthest pair of vertices) is less than or equal to $N/2^{m+1}$

Proof. Assume on the contrary that G^0 is not connected. Let $\{\mathbf{x}, \mathbf{x}'\}$ be a pair of vertices which minimizes $\|\mathbf{x} - \mathbf{x}'\|_1$ subject to the condition that there does not exist any path between \mathbf{x} and \mathbf{x}' . Without loss of generality, we can assume that $\exists j' \in J$, $x(\mathbf{2}; j') < x'(\mathbf{2}; j')$, where $\mathbf{2}$ is the all-two vector in \mathbb{B}^m . It directly implies the followings;

1. $x(\mathbf{i}; j') < x'(\mathbf{i}; j')$ for any $\mathbf{i} \in \mathbb{B}^m$ satisfying $p(\mathbf{i}) = p(\mathbf{2})$,
2. $x(\mathbf{i}; j') > x'(\mathbf{i}; j')$ for any $\mathbf{i} \in \mathbb{B}^m$ satisfying $p(\mathbf{i}) \neq p(\mathbf{2})$,
3. $|x(\mathbf{i}; j') - x'(\mathbf{i}; j')| = |x(\mathbf{2}; j') - x'(\mathbf{2}; j')|$ for any $\mathbf{i} \in \mathbb{B}^m$.

Since $\sum_{j \in J} x(\mathbf{2}; j) = \sum_{j \in J} x'(\mathbf{2}; j)$, there exists a column index j'' satisfying $x(\mathbf{2}; j'') > x'(\mathbf{2}; j'')$. Then we have the following properties;

1. $x(\mathbf{i}; j'') > x'(\mathbf{i}; j'')$ for any $\mathbf{i} \in \mathbb{B}^m$ satisfying $p(\mathbf{i}) = p(\mathbf{2})$,
2. $x(\mathbf{i}; j'') < x'(\mathbf{i}; j'')$ for any $\mathbf{i} \in \mathbb{B}^m$ satisfying $p(\mathbf{i}) \neq p(\mathbf{2})$,
3. $|x(\mathbf{i}; j'') - x'(\mathbf{i}; j'')| = |x(\mathbf{2}; j'') - x'(\mathbf{2}; j'')|$ for any $\mathbf{i} \in \mathbb{B}^m$.

The vector $\mathbf{x}'' = \mathbf{x} + \Delta[j', j'']$ is non-negative and so $\mathbf{x}'' \in \mathcal{T}$. Since $\mathbf{x}'' \in \mathbb{N}^0(\mathbf{x})$, there does not exist any path between \mathbf{x}'' and \mathbf{x}' . The inequality $\|\mathbf{x} - \mathbf{x}'\|_1 > \|\mathbf{x}'' - \mathbf{x}'\|_1$ contradicts with the minimality of $\|\mathbf{x} - \mathbf{x}'\|_1$. The definition of \mathbf{x}'' implies that $\|\mathbf{x} - \mathbf{x}''\|_1 = 2^{m+1}$.

The above procedure decreases the distance between a distinct pair of vertices and the decrement is 2^{m+1} . If we apply the procedure $\lfloor \|\mathbf{x} - \mathbf{x}'\|_1 / 2^{m+1} \rfloor$ times, the distance of two vertices is less than 2^{m+1} . It implies that the obtained pair of vertices are identical. Thus the diameter of G^0 is less than or equal to $N/2^{m+1}$. \square

The above lemma indicates the existence of an irreducible Markov chain on \mathcal{T} such that the transition probability of an ordered pair of tables $(\mathbf{x}, \mathbf{x}')$ is positive if and only if \mathbf{x} and \mathbf{x}' are adjacent on G^0 . When $m = 1$, this chain is a special case of the Markov chain proposed by Diaconis and Saloff-Coste [10]. However, as discussed in Dyer and Greenhill [12], the mixing rate of Diaconis and Saloff-Coste's chain is low. In the following, we describe our chain, which is an extension of the chain discussed by Dyer and Greenhill [12] for contingency tables with two rows.

For any table $\mathbf{x} \in \mathcal{T}$ and any pair of distinct column indices $\{j', j''\}$, we define the following set of tables;

$$\mathcal{N}(\mathbf{x}, \{j', j''\}) \stackrel{\text{def.}}{=} \left\{ (\mathbf{y}(j'), \mathbf{y}(j'')) \in \mathbb{Z}_+^{\mathbb{B}^m \times \{j', j''\}} \mid \exists \theta \in \mathbb{Z}, (\mathbf{y}(j'), \mathbf{y}(j'')) = (\mathbf{x}(j'), \mathbf{x}(j'')) + \theta(\Delta, -\Delta) \geq \mathbf{0} \right\}.$$

Markov Chain \mathcal{M}^1

We introduce our chain \mathcal{M}^1 with state space \mathcal{T} . For any table $\mathbf{x} \in \mathcal{T}$ and any pair of distinct column indices $\{j', j''\}$, we define the following set of tables;

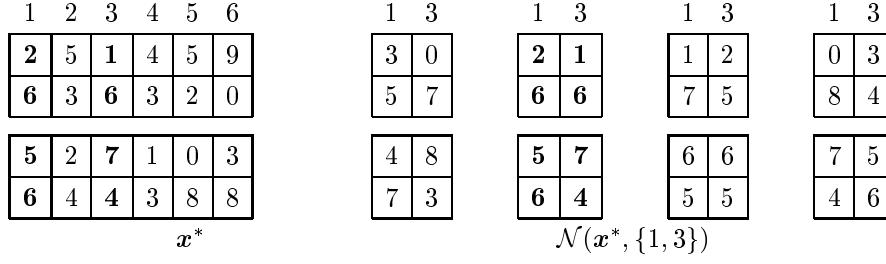


Figure 3: Vectors in $\mathcal{N}(\mathbf{x}^*, \{1, 3\})$.

$$\begin{aligned} \mathcal{N}^1(\mathbf{x}, \{j', j''\}) &\stackrel{\text{def.}}{=} \{\mathbf{x}' \in \mathcal{T} \mid \mathbf{x}'(j) = \mathbf{x}(j) \ (\forall j \in J \setminus \{j', j''\}), (\mathbf{x}'(j'), \mathbf{x}'(j'')) \in \mathcal{N}(\mathbf{x}, \{j', j''\})\} \\ &= \{\mathbf{x}' \in \mathcal{T} \mid \exists \theta \in \mathbb{Z}, \mathbf{x}' = \mathbf{x} + \theta \Delta[j', j''] \geq \mathbf{0}\}. \end{aligned}$$

The Markov chain \mathcal{M}^1 with the state space \mathcal{T} is defined by the following transition procedure. We denote the state of the chain \mathcal{M}^1 at time t by X_t and the element of X_t indexed by $(i; j)$ is denoted by $X_t(i; j)$. Then the state X_{t+1} at time $t+1$ is determined as follows. First, choose a pair of distinct column indices $\{j', j''\}$ randomly. Next, choose a table X_{t+1} from $\mathcal{N}^1(X_t, \{j', j''\})$ at random.

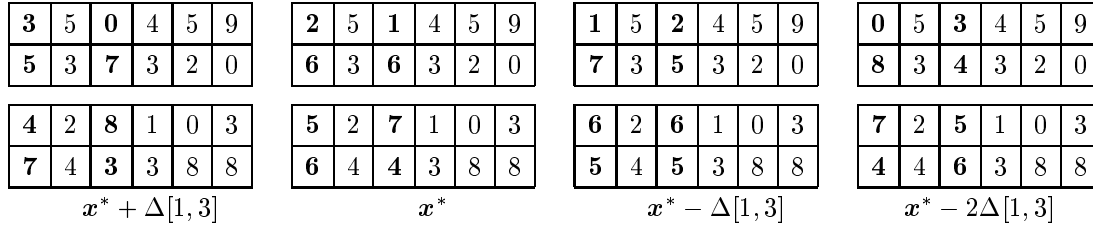


Figure 4: The set of neighbors $\mathcal{N}^1(\mathbf{x}^*, \{1, 3\})$.

We estimate the mixing time of our chain \mathcal{M}^1 . According to the definition, it is clear that \mathcal{M}^1 is aperiodic and irreducible. The transition probability of \mathcal{M}^1 from \mathbf{x} to \mathbf{y} , denoted by $P_{\mathcal{M}^1}(\mathbf{x}, \mathbf{y})$ is

$$P_{\mathcal{M}^1}(\mathbf{x}, \mathbf{y}) = \begin{cases} \left(\binom{n}{2} |\mathcal{N}(\mathbf{x}, \{j', j''\})| \right)^{-1} & (\text{if } \mathbf{y} \in \mathcal{N}^1(\mathbf{x}, \{j', j''\})), \\ \sum_{j' < j''} \left(\binom{n}{2} |\mathcal{N}(\mathbf{x}, \{j', j''\})| \right)^{-1} & (\mathbf{x} = \mathbf{y}), \\ 0 & (\text{otherwise}). \end{cases}$$

Since $P_{\mathcal{M}^1}(\mathbf{x}, \mathbf{y}) = P_{\mathcal{M}^1}(\mathbf{y}, \mathbf{x})$, the stationary distribution of the chain is uniform.

First, we introduce a directed graph G^1 with the vertex set \mathcal{T} and the arc set $A = \{(\mathbf{x}, \mathbf{x}') \mid \mathbf{x}' \in \mathcal{N}^0(\mathbf{x})\}$. We define that the length $\ell(a)$ of each arc $a \in A$ is equal to 1. The distance of any ordered pair of vertices $(\mathbf{x}, \mathbf{x}')$ on G^1 is denoted by $\ell(\mathbf{x}, \mathbf{x}')$. Next, we define a joint process $(X, Y) \mapsto (X', Y')$ with respect to \mathcal{M}^1 . For any pair of tables $(X, Y) \in A$, we define the transition probability of our joint process from (X, Y) to (X', Y') . Without loss of generality, we can assume that $X(1) \neq Y(1)$, $X(2) \neq Y(2)$ and $X(j) = Y(j)$ for all $j \in J \setminus \{1, 2\}$. In the joint process, we choose a pair of distinct column indices (j', j'') .

Case 1: When $\{j', j''\} \subseteq \{3, \dots, n\}$, it is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$ and so we choose a pair $(Z(j'), Z(j''))$ from $\mathcal{N}(X, \{j', j''\})$ at random. We set X' and Y' to the contingency table obtained from X and Y by replacing $(X(j'), X(j''))$ and $(Y(j'), Y(j''))$ by $(Z(j'), Z(j''))$, respectively. Then, it is clear that (X', Y') is also in A and so $\ell(X', Y') = 1$.

Case 2: Next, consider the case that $\{j', j''\} = \{1, 2\}$. It is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$. We construct X' and Y' by using the same manner of Case 1. Then, we have $X' = Y'$ and $\ell(X', Y') = 0$.

Case 3: Finally, we consider the case that $j' = 1$ and $j'' = 3$. Other cases are treated in the same way as follows.

Case 3-1: Consider the case that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})|$.

We denote $N^1(X, \{j', j''\}) = \{X^1, X^2, \dots, X^k\}$ and $N^1(Y, \{j', j''\}) = \{Y^1, Y^2, \dots, Y^k\}$. By arranging the order of the elements, we assume that $X^1(\mathbf{1}; 1) > X^2(\mathbf{1}; 1) > \dots > X^k(\mathbf{1}; 1)$ and $Y^1(\mathbf{1}; 1) > Y^2(\mathbf{1}; 1) > \dots > Y^k(\mathbf{1}; 1)$. Then we choose (X', Y') randomly from $\{(X^1, Y^1), (X^2, Y^2), \dots, (X^k, Y^k)\}$. It is easy to see that $(X', Y') \in A$ and so $\ell(X', Y') = 1$.

Case 3-2: We only need to consider the case that $|\mathcal{N}(X, \{j', j''\})| > |\mathcal{N}(Y, \{j', j''\})|$ without loss of generality. Since $(X, Y) \in A$, it is easy to show that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})| + 1$. By arranging the order of elements in $N^1(X, \{j', j''\}) = \{X^1, X^2, \dots, X^{k+1}\}$ and $N^1(Y, \{j', j''\}) = \{Y^1, Y^2, \dots, Y^k\}$, we can assume that $X^1(\mathbf{1}; 1) > X^2(\mathbf{1}; 1) > \dots > X^{k+1}(\mathbf{1}; 1)$ and $Y^1(\mathbf{1}; 1) > Y^2(\mathbf{1}; 1) > \dots > Y^k(\mathbf{1}; 1)$. Then we choose (X', Y') as follows;

$$(X', Y') = \begin{cases} (X^i, Y^i) & \text{with probability } (k-i+1)/k(k+1) & \text{for } i \in \{1, 2, \dots, k\}, \\ (X^{i+1}, Y^i) & \text{with probability } i/k(k+1) & \text{for } i \in \{1, 2, \dots, k\}, \end{cases}$$

where the sum total of the probabilities is $(1+2+\dots+k)/k(k+1) + (k+\dots+2+1)/k(k+1) = 1$. Figure 5 shows an example. Clearly from the definition, $(X^i, Y^i), (X^i, Y^{i+1}) \in A$ for each $i \in \{1, 2, \dots, k\}$ and so $\ell(X', Y') = 1$.

From the above, we have

$$\mathbb{E}[\ell(X', Y')] = \left(1 - \binom{n}{2}^{-1}\right).$$

It implies the following result.

Theorem 1 *The Markov chain \mathcal{M}^1 has the mixing time $\tau_1(\varepsilon)$ satisfying that*

$$\tau_1(\varepsilon) \leq (1/2)n(n-1) \ln(dn/(2\varepsilon)),$$

where d is the average of the values in cells, i.e., $d = N/(2^m n)$.

Proof. The diameter of the graph G^1 is equal to that of G^0 and so less than or equal to $N/2^{m+1}$. Path coupling lemma induces the desired result. \square

4 Markov Chain for Conditional Multinomial Distribution

In this section, we consider a conditional multinomial distribution given marginal sum. The distribution function $\psi : \mathcal{T} \rightarrow [0, 1]$ is defined by

$$\psi(X) \stackrel{\text{def.}}{=} (1/\rho(\mathcal{T})) \prod_{(\mathbf{i}; j) \in \mathbb{B}^m \times J} (X(\mathbf{i}; j))^{-1}, \quad \text{where} \quad \rho(\mathcal{T}) \stackrel{\text{def.}}{=} \sum_{\mathbf{y} \in \mathcal{T}} \prod_{(\mathbf{i}; j) \in \mathbb{B}^m \times J} (\mathbf{y}(\mathbf{i}; j))^{-1}.$$

3	5	0	4	5	9
5	3	7	3	2	0

$$X^1 = X + \Delta[1, 3]$$

$$= Y^1 + \Delta[1, 2]$$

2	5	1	4	5	9
6	3	6	3	2	0

$$X^2 = X$$

$$= Y^2 + \Delta[1, 2]$$

1	5	2	4	5	9
7	3	5	3	2	0

$$X^3 = X - \Delta[1, 3]$$

$$= Y^3 + \Delta[1, 2]$$

0	5	3	4	5	9
8	3	4	3	2	0

$$X^4 = X - 2\Delta[1, 3]$$

2	6	0	4	5	9
6	2	7	3	2	0

$$Y^1 = Y + \Delta[1, 3]$$

$$= X^2 + \Delta[2, 3]$$

1	6	1	4	5	9
7	2	6	3	2	0

$$Y^2 = Y = X + \Delta[2, 1]$$

$$= X^3 + \Delta[2, 3]$$

0	6	2	4	5	9
8	2	5	3	2	0

$$Y^3 = Y - \Delta[1, 3]$$

$$= X^4 + \Delta[2, 3]$$

$$\Pr[(X', Y') = (X^1, Y^1) | (j', j'') = (1, 3)] = 3/12, \quad \Pr[(X', Y') = (X^2, Y^1) | (j', j'') = (1, 3)] = 1/12,$$

$$\Pr[(X', Y') = (X^2, Y^2) | (j', j'') = (1, 3)] = 2/12, \quad \Pr[(X', Y') = (X^3, Y^2) | (j', j'') = (1, 3)] = 2/12,$$

$$\Pr[(X', Y') = (X^3, Y^3) | (j', j'') = (1, 3)] = 1/12, \quad \Pr[(X', Y') = (X^4, Y^3) | (j', j'') = (1, 3)] = 3/12.$$

Figure 5: An example of Case 3-2.

Next, we describe a Markov chain which has a conditional multinomial distribution as a stationary distribution.

Markov Chain \mathcal{M}^2

First, we introduce a distribution function $\psi(\mathbf{x}, \{j', j''\})$ defined on $N^1(\mathbf{x}, \{j', j''\})$ as follows

$$\psi(\mathbf{x}, \{j', j''\}) : Y \mapsto (1/\rho(\mathbf{x}; \{j', j''\})) \prod_{(\mathbf{i}; j) \in B \times \{j', j''\}} (Y(\mathbf{i}; j))^{-1},$$

where

$$\rho(\mathbf{x}; \{j', j''\}) \stackrel{\text{def.}}{=} \sum_{\mathbf{y} \in N^1(\mathbf{x}, \{j', j''\})} \prod_{(\mathbf{i}; j) \in B \times \{j', j''\}} (\mathbf{y}(\mathbf{i}; j))^{-1}.$$

The Markov chain \mathcal{M}^2 with the state space \mathcal{T} is defined by the following transition procedure. We denote the state of the chain \mathcal{M}^2 at time t by X_t . Then the state X_{t+1} at time $t+1$ is determined as follows. First, choose a pair of distinct column indices $\{j', j''\}$ randomly. Next, choose a table X_{t+1} from $N^1(X_t, \{j', j''\})$ under the distribution function $\psi(X_t, \{j', j''\})$.

According to the definition, it is clear that \mathcal{M}^2 is aperiodic and irreducible. Since the detailed balance equations hold, the function ψ is a unique stationary distribution function of \mathcal{M}^2 .

Next, we define a joint process $(X, Y) \mapsto (X', Y')$ with respect to \mathcal{M}^2 . Recall that A is the arc set of digraph G^1 defined in the previous section. For any pair of tables $(X, Y) \in A$, we define the transition probability of our joint process from (X, Y) to (X', Y') . In the following, we consider the case that $(X, Y) \in A$. Without loss of generality, we can assume that $X(1) \neq Y(1)$, $X(2) \neq Y(2)$ and $X(j) = Y(j)$ for all $j \in J \setminus \{1, 2\}$. In the joint process, we choose a pair of distinct column indices (j', j'') .

Case 1: When $\{j', j''\} \subseteq \{3, \dots, n\}$, it is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$ and so we choose a pair $(Z(j'), Z(j''))$ from $\mathcal{N}(X, \{j', j''\})$ under the distribution function $\psi(X, \{j', j''\})$. We set X' and

Y' to the contingency table obtained from X and Y by replacing $(X(j'), X(j''))$ and $(Y(j'), Y(j''))$ by $(Z(j'), Z(j''))$, respectively. Then, it is clear that (X', Y') is also in A and so $\ell(X', Y') = 1$.

Case 2: Next, consider the case that $\{j', j''\} = \{1, 2\}$. It is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$. We construct X' and Y' by using the same manner of Case 1. Then, we have $X' = Y'$ and $\ell(X', Y') = 0$.

Case 3: Finally, we consider the case that $j' = 1$ and $j'' = 3$. Other cases are treated in a similar way.

Case 3-1: Consider the case that $|\mathcal{N}(X, \{j', j''\})| \neq |\mathcal{N}(Y, \{j', j''\})|$. We can assume that $|\mathcal{N}(X, \{j', j''\})| > |\mathcal{N}(Y, \{j', j''\})|$ without loss of generality. Since $(X, Y) \in A$, it is easy to show that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})| + 1$. By arranging the order of elements in $\mathcal{N}^1(X, \{j', j''\}) = \{X^1, X^2, \dots, X^{k+1}\}$ and $\mathcal{N}^1(Y, \{j', j''\}) = \{Y^1, Y^2, \dots, Y^k\}$, we can assume that $X^1(\mathbf{1}; 1) > X^2(\mathbf{1}; 1) > \dots > X^{k+1}(\mathbf{1}; 1)$ and $Y^1(\mathbf{1}; 1) > Y^2(\mathbf{1}; 1) > \dots > Y^k(\mathbf{1}; 1)$.

Exactly one of the following two cases holds;

- (i) $(Y^1(1), Y^2(1), \dots, Y^k(1)) = (X^1(1), X^2(1), \dots, X^k(1))$ and $(Y^1(3), Y^2(3), \dots, Y^k(3)) = (X^2(3), X^3(3), \dots, X^{k+1}(3))$,
- (ii) $(Y^1(3), Y^2(3), \dots, Y^k(3)) = (X^1(3), X^2(3), \dots, X^k(3))$ and $(Y^1(1), Y^2(1), \dots, Y^k(1)) = (X^2(1), X^3(1), \dots, X^{k+1}(1))$.

In the following, we consider Case (i). We can deal with Case (ii) in a similar way.

We choose (X', Y') as follows;

$$(X', Y') = \begin{cases} (X^1, Y^1) & \text{with probability } \psi(X, \{j', j''\})(X^1), \\ (X^i, Y^i) & \text{with probability } \sum_{i'=1}^i \psi(X, \{j', j''\})(X^{i'}) - \sum_{i'=1}^{i-1} \psi(Y, \{j', j''\})(Y^{i'}) \\ & \text{for } i \in \{1, 2, \dots, k\}, \\ (X^{i+1}, Y^i) & \text{with probability } \sum_{i'=1}^i \psi(Y, \{j', j''\})(Y^{i'}) - \sum_{i'=1}^i \psi(X, \{j', j''\})(X^{i'}) \\ & \text{for } i \in \{1, 2, \dots, k\}. \end{cases}$$

It is clear that the probabilities described above satisfy the equalities appearing in the definition of joint process. We need to show the non-negativity of the above probabilities. From the definition, $(X^i, Y^i), (X^i, Y^{i+1}) \in A$ for each $i \in \{1, 2, \dots, k\}$ and so $\ell(X', Y') = 1$. To show the non-negativity, we need the following Lemma.

Lemma 2 *Let $(\alpha_1, \alpha_2, \dots, \alpha_{k+1})$ and $(\beta_1, \dots, \beta_k)$ be a pair of positive vectors satisfying that*

$$\frac{\alpha_1}{\beta_1} \leq \frac{\alpha_2}{\beta_2} \leq \dots \leq \frac{\alpha_k}{\beta_k} \quad \text{and} \quad \frac{\beta_1}{\alpha_2} \leq \frac{\beta_2}{\alpha_3} \leq \dots \leq \frac{\beta_k}{\alpha_{k+1}}.$$

Let F be an index set defined by $F \stackrel{\text{def}}{=} \{(1, 1), (2, 2), \dots, (k, k), (2, 1), (3, 2), \dots, (k+1, k)\}$ and $\gamma \in \mathbf{R}^F$ be a vector defined by;

$$\begin{aligned} \gamma(1, 1) &= \alpha_1/A, \\ \gamma(i, i) &= (\alpha_1 + \dots + \alpha_i)/A - (\beta_1 + \dots + \beta_{i-1})/B \quad (i = 1, 2, \dots, k), \\ \gamma(i+1, i) &= (\beta_1 + \dots + \beta_i)/B - (\alpha_1 + \dots + \alpha_i)/A \quad (i = 1, \dots, k), \end{aligned}$$

where $A = \alpha_1 + \dots + \alpha_{k+1}$ and $B = \beta_1 + \dots + \beta_k$. Then the vector γ is non-negative.

Proof.

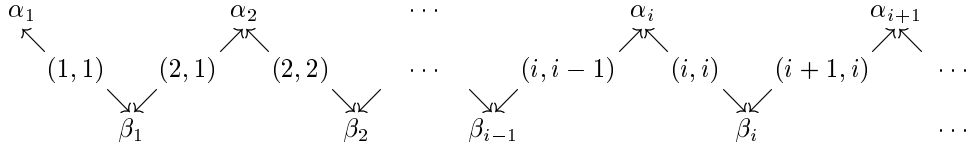


Figure 6: Index set F .

(0) Clearly, $\gamma(1, 1) \geq 0$.

(1) Let $i^* \stackrel{\text{def.}}{=} \max\{i \mid \beta_i/\alpha_{i+1} \leq B/A\}$. In the case that $\{i \mid \beta_i/\alpha_{i+1} \leq B/A\} = \emptyset$, we set $i^* \stackrel{\text{def.}}{=} -\infty$.

From the definition, if $i^* \neq -\infty$, then we have

$$\frac{\beta_1}{\alpha_2} \leq \frac{\beta_2}{\alpha_3} \leq \dots \leq \frac{\beta_{i^*}}{\alpha_{i^*+1}} \leq \frac{B}{A} < \frac{\beta_{i^*+1}}{\alpha_{i^*+2}} \leq \dots \leq \frac{\beta_k}{\alpha_{k+1}}.$$

For any index $i \in (-\infty, i^*] \cap \{1, 2, \dots, k\}$, $\gamma(i, i)$ satisfies

$$\begin{aligned} \gamma(i, i) &= (\alpha_1 + \dots + \alpha_i)/A - (\beta_1 + \dots + \beta_{i-1})/B \\ &= \alpha_1/A + (\alpha_2/A - \beta_1/B) + \dots + (\alpha_i/A - \beta_{i-1}/B) \\ &\geq (\alpha_2/B)(B/A - \beta_1/\alpha_2) + \dots + (\alpha_i/B)(B/A - \beta_{i-1}/\alpha_i) \geq 0. \end{aligned}$$

For any index $i \in [i^* + 1, +\infty) \cap \{1, 2, \dots, k\}$, the following inequality holds;

$$\begin{aligned} \gamma(i, i) &= (\alpha_1 + \dots + \alpha_i)/A - (\beta_1 + \dots + \beta_{i-1})/B \\ &= (A - (\alpha_{i+1} + \dots + \alpha_{k+1}))/A - (B - (\beta_i + \dots + \beta_k))/B \\ &= (\beta_i/B - \alpha_{i+1}/A) + \dots + (\beta_k/B - \alpha_{k+1}/A) \\ &\geq (\alpha_{i+1}/B)(\beta_i/\alpha_{i+1} - B/A) + \dots + (\alpha_{k+1}/B)(\beta_k/\alpha_{k+1} - B/A) \geq 0. \end{aligned}$$

(2) Let i^* be the index defined by $i^* \stackrel{\text{def.}}{=} \max\{i \mid \alpha_i/\beta_i \leq A/B\}$. If $\{i \mid \alpha_i/\beta_i \leq A/B\} = \emptyset$, we define $i^* = -\infty$. Thus, in the case that $i^* \neq -\infty$, following inequalities hold;

$$\frac{\alpha_1}{\beta_1} \leq \frac{\alpha_2}{\beta_2} \leq \dots \leq \frac{\alpha_{i^*}}{\beta_{i^*}} \leq \frac{A}{B} < \frac{\alpha_{i^*+1}}{\beta_{i^*+1}} \leq \dots \leq \frac{\alpha_k}{\beta_k}.$$

For any index $i \in (-\infty, i^*] \cap \{1, 2, \dots, k\}$, $\gamma(i+1, i)$ satisfies

$$\begin{aligned} \gamma(i+1, i) &= (\beta_1 + \dots + \beta_i)/B - (\alpha_1 + \dots + \alpha_i)/A \\ &= (\beta_1/B - \alpha_1/A) + \dots + (\beta_i/B - \alpha_i/A) \\ &= (\beta_1/A)(A/B - \alpha_1/\beta_1) + \dots + (\beta_i/A)(A/B - \alpha_i/\beta_i) \geq 0. \end{aligned}$$

For any index $i \in [i^* + 1, +\infty) \cap \{1, 2, \dots, k\}$, we can show the non-negativity of $\gamma(i+1, i)$ as follows;

$$\begin{aligned} \gamma(i+1, i) &= (\beta_1 + \dots + \beta_i)/B - (\alpha_1 + \dots + \alpha_i)/A \\ &= (B - (\beta_{i+1} + \dots + \beta_k))/B - (A - (\alpha_{i+1} + \dots + \alpha_{k+1}))/A \\ &= (\alpha_{i+1}/A - \beta_{i+1}/B) + \dots + (\alpha_k/A - \beta_k/B) + \alpha_{k+1}/A \\ &\geq (\beta_{i+1}/A)(\alpha_{i+1}/\beta_{i+1} - A/B) + \dots + (\beta_k/A)(\alpha_k/\beta_k - A/B) \geq 0. \end{aligned}$$

From the above, we have the desired result. \square

Let $B_{\mathbf{1}}^m \stackrel{\text{def.}}{=} \{\mathbf{i} \in B^m \mid p(\mathbf{i}) = p(\mathbf{1})\}$. For any index $\mathbf{i} \in B_{\mathbf{1}}^m$, $(X^1(\mathbf{i}; 1), X^2(\mathbf{i}; 1), \dots, X^{k+1}(\mathbf{i}; 1))$ is an arithmetic sequence of non-negative integers with the common difference -1 . For each index $\mathbf{i} \in B^m$, we denote the index obtained from \mathbf{i} by flipping the m th element (i.e., the last element) by $\hat{\mathbf{i}}$. Then, for any index $\mathbf{i} \in B_{\mathbf{1}}^m$, the sequence $(X^1(\hat{\mathbf{i}}; 1), X^2(\hat{\mathbf{i}}; 1), \dots, X^{k+1}(\hat{\mathbf{i}}; 1))$ is an arithmetic sequence of non-negative integers with the common difference 1. It is also easy to show that for any index $\mathbf{i} \in B_{\mathbf{1}}^m$, the sequences $(X^1(\mathbf{i}; 3), X^2(\mathbf{i}; 3), \dots, X^{k+1}(\mathbf{i}; 3))$ and $(X^1(\hat{\mathbf{i}}; 3), X^2(\hat{\mathbf{i}}; 3), \dots, X^{k+1}(\hat{\mathbf{i}}; 3))$ are arithmetic sequences of non-negative integers with the common differences 1 and -1 , respectively. The above properties imply the following lemma.

Lemma 3 *Assume that the condition ;*

$$(i) \quad (Y^1(1), Y^2(1), \dots, Y^k(1)) = (X^1(1), X^2(1), \dots, X^k(1)) \text{ and} \\ (Y^1(3), Y^2(3), \dots, Y^k(3)) = (X^2(3), X^3(3), \dots, X^{k+1}(3))$$

is satisfied. Then the following inequalities hold;

$$\frac{\psi(X, \{j', j''\})(X^1)}{\psi(Y, \{j', j''\})(Y^1)} \leq \frac{\psi(X, \{j', j''\})(X^2)}{\psi(Y, \{j', j''\})(Y^2)} \leq \dots \leq \frac{\psi(X, \{j', j''\})(X^k)}{\psi(Y, \{j', j''\})(Y^k)}, \\ \frac{\psi(Y, \{j', j''\})(Y^1)}{\psi(X, \{j', j''\})(X^2)} \leq \frac{\psi(Y, \{j', j''\})(Y^2)}{\psi(X, \{j', j''\})(X^3)} \leq \dots \leq \frac{\psi(Y, \{j', j''\})(Y^k)}{\psi(X, \{j', j''\})(X^{k+1})}.$$

Proof. We introduce some notations for simplicity. For any index $\mathbf{i} \in B_{\mathbf{1}}^m$, we define

$$(\eta_{\mathbf{i}}^1, \eta_{\mathbf{i}}^2, \dots, \eta_{\mathbf{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\mathbf{i}; 1), X^2(\mathbf{i}; 1), \dots, X^{k+1}(\mathbf{i}; 1)), \\ (\bar{\eta}_{\mathbf{i}}^1, \bar{\eta}_{\mathbf{i}}^2, \dots, \bar{\eta}_{\mathbf{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\hat{\mathbf{i}}; 3), X^2(\hat{\mathbf{i}}; 3), \dots, X^{k+1}(\hat{\mathbf{i}}; 3)), \\ (\xi_{\mathbf{i}}^1, \xi_{\mathbf{i}}^2, \dots, \xi_{\mathbf{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\mathbf{i}; 3), X^2(\mathbf{i}; 3), \dots, X^{k+1}(\mathbf{i}; 3)), \\ (\bar{\xi}_{\mathbf{i}}^1, \bar{\xi}_{\mathbf{i}}^2, \dots, \bar{\xi}_{\mathbf{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\hat{\mathbf{i}}; 1), X^2(\hat{\mathbf{i}}; 1), \dots, X^{k+1}(\hat{\mathbf{i}}; 1)).$$

Then, both $(\eta_{\mathbf{i}}^1, \eta_{\mathbf{i}}^2, \dots, \eta_{\mathbf{i}}^{k+1})$ and $(\bar{\eta}_{\mathbf{i}}^1, \bar{\eta}_{\mathbf{i}}^2, \dots, \bar{\eta}_{\mathbf{i}}^{k+1})$ are arithmetic sequences of non-negative integers with common difference -1 . Both of the sequences $(\xi_{\mathbf{i}}^1, \xi_{\mathbf{i}}^2, \dots, \xi_{\mathbf{i}}^{k+1})$ and $(\bar{\xi}_{\mathbf{i}}^1, \bar{\xi}_{\mathbf{i}}^2, \dots, \bar{\xi}_{\mathbf{i}}^{k+1})$ are arithmetic sequences of non-negative integers with common difference 1.

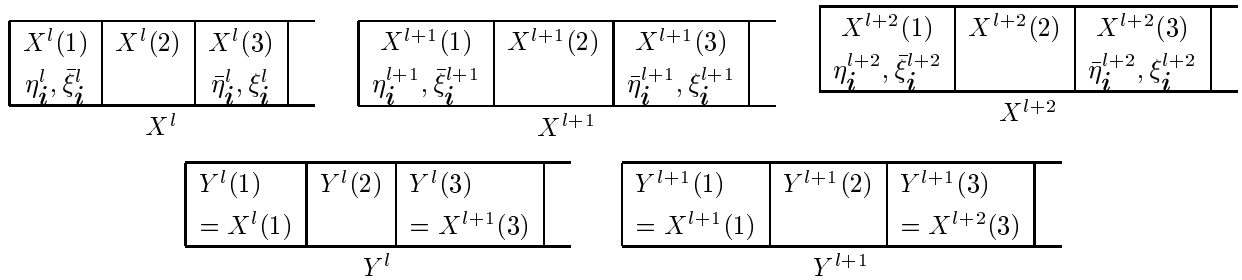


Figure 7: Condition (i) of Lemma 3.

From the above, it is easy to see that

$$\left(\frac{\psi(X, \{j', j''\})(X^l)}{\psi(Y, \{j', j''\})(Y^l)} \right) \left(\frac{\psi(X, \{j', j''\})(X^{l+1})}{\psi(Y, \{j', j''\})(Y^{l+1})} \right)^{-1}$$

$$\begin{aligned}
&= \left(\left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^l! \bar{\eta}_i^l! \xi_i^l! \bar{\xi}_i^l!)^{-1} \right) \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+1}! \bar{\eta}_i^{l+1}! \xi_i^{l+1}! \bar{\xi}_i^{l+1}!)^{-1} \right)^{-1} \right) \\
&\quad \left(\left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+1}! \bar{\eta}_i^{l+1}! \xi_i^{l+1}! \bar{\xi}_i^{l+1}!)^{-1} \right) \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+1}! \bar{\eta}_i^{l+2}! \xi_i^{l+2}! \bar{\xi}_i^{l+1}!)^{-1} \right)^{-1} \right)^{-1} \\
&= \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\bar{\eta}_i^l)^{-1} \xi_i^{l+1} \right) \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\bar{\eta}_i^{l+1})^{-1} \xi_i^{l+2} \right) \left[\begin{array}{l} \text{since } \bar{\eta}_i^l > \bar{\eta}_i^{l+1} > \bar{\eta}_i^{l+2} \\ \text{and } \xi_i^l < \xi_i^{l+1} < \xi_i^{l+2} \end{array} \right] \\
&= \prod_{\mathbf{i} \in \mathbb{B}_1^m} \frac{\xi_i^{l+1} \bar{\eta}_i^{l+1}}{\bar{\eta}_i^l \xi_i^{l+2}} = \prod_{\mathbf{i} \in \mathbb{B}_1^m} \frac{\bar{\eta}_i^{l+1} \xi_i^{l+1}}{(\bar{\eta}_i^{l+1} + 1)(\xi_i^{l+1} + 1)} \leq 1.
\end{aligned}$$

In a similar way, we can show that

$$\begin{aligned}
&\left(\frac{\psi(Y, \{j', j''\})(Y^l)}{\psi(X, \{j', j''\})(X^{l+1})} \right) \left(\frac{\psi(Y, \{j', j''\})(Y^{l+1})}{\psi(X, \{j', j''\})(X^{l+2})} \right)^{-1} \\
&= \left(\left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^l! \bar{\eta}_i^{l+1}! \xi_i^{l+1}! \bar{\xi}_i^l!)^{-1} \right) \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+1}! \bar{\eta}_i^{l+1}! \xi_i^{l+1}! \bar{\xi}_i^{l+1}!)^{-1} \right)^{-1} \right) \\
&\quad \left(\left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+1}! \bar{\eta}_i^{l+2}! \xi_i^{l+2}! \bar{\xi}_i^{l+1}!)^{-1} \right) \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+2}! \bar{\eta}_i^{l+2}! \xi_i^{l+2}! \bar{\xi}_i^{l+2}!)^{-1} \right)^{-1} \right)^{-1} \\
&= \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^l)^{-1} \bar{\xi}_i^{l+1} \right) \left(\prod_{\mathbf{i} \in \mathbb{B}_1^m} (\eta_i^{l+1})^{-1} \bar{\xi}_i^{l+2} \right)^{-1} = \prod_{\mathbf{i} \in \mathbb{B}_1^m} \frac{\bar{\xi}_i^{l+1} \eta_i^{l+1}}{\eta_i^l \bar{\xi}_i^{l+2}} = \prod_{\mathbf{i} \in \mathbb{B}_1^m} \frac{\eta_i^{l+1} \bar{\xi}_i^{l+1}}{(\eta_i^{l+1} + 1)(\bar{\xi}_i^{l+1} + 1)} \leq 1.
\end{aligned}$$

□

If we set $\psi(X, \{j', j''\})(X^\ell) = \alpha_\ell$ and $\psi(Y, \{j', j''\})(Y^\ell) = \beta_\ell$, above lemmas directly imply the non-negativity of the transition probabilities of our joint process.

Case 3-2: Consider the case that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})|$. We denote $\mathbb{N}^1(X, \{j', j''\}) = \{X^1, X^2, \dots, X^k\}$ and $\mathbb{N}^1(Y, \{j', j''\}) = \{Y^1, Y^2, \dots, Y^k\}$. By arranging the order of the elements, we can assume that $X^1(\mathbf{1}; 1) > X^2(\mathbf{1}; 1) > \dots > X^k(\mathbf{1}; 1)$ and $Y^1(\mathbf{1}; 1) > Y^2(\mathbf{1}; 1) > \dots > Y^k(\mathbf{1}; 1)$. Exactly one of the following four cases holds;

- (i) $(X^1(1), X^2(1), \dots, X^k(1)) = (Y^1(1), Y^2(1), \dots, Y^k(1))$ and $(X^2(3), X^3(3), \dots, X^k(3)) = (Y^1(3), Y^2(3), \dots, Y^{k-1}(3))$,
- (ii) $(X^1(3), X^2(3), \dots, X^k(3)) = (Y^1(3), Y^2(3), \dots, Y^k(3))$ and $(X^2(1), X^3(1), \dots, X^k(1)) = (Y^1(1), Y^2(1), \dots, Y^{k-1}(1))$,
- (iii) $(X^1(1), X^2(1), \dots, X^k(1)) = (Y^1(1), Y^2(1), \dots, Y^k(1))$ and $(X^1(3), X^2(3), \dots, X^{k-1}(3)) = (Y^2(3), Y^3(3), \dots, Y^k(3))$,
- (iv) $(X^1(3), X^2(3), \dots, X^k(3)) = (Y^1(3), Y^2(3), \dots, Y^k(3))$ and $(X^1(1), X^2(1), \dots, X^{k-1}(1)) = (Y^2(1), Y^3(1), \dots, Y^k(1))$.

In the following, we consider Case (i). We can deal with Cases (ii),(iii),(iv) in a similar way.

We choose (X', Y') as follows;

$$(X', Y') = \begin{cases} (X^1, Y^1) & \text{with probability } \psi(X, \{j', j''\})(X^1), \\ (X^i, Y^i) & \text{with probability } \sum_{i'=1}^i \psi(X, \{j', j''\})(X^{i'}) - \sum_{i'=1}^{i-1} \psi(Y, \{j', j''\})(Y^{i'}) \\ & \text{for } i \in \{2, \dots, k\}, \\ (X^{i+1}, Y^i) & \text{with probability } \sum_{i'=1}^i \psi(Y, \{j', j''\})(Y^{i'}) - \sum_{i'=1}^i \psi(X, \{j', j''\})(X^{i'}) \\ & \text{for } i \in \{1, 2, \dots, k-1\}, \end{cases}$$

We need to show the non-negativity of the probabilities defined above. Clearly from the definition, $(X^i, Y^i), (X^{i+1}, Y^i) \in A$ for each $i \in \{1, 2, \dots, k\}$ and so $\ell(X', Y') = 1$. To show the non-negativity, we need the following lemma.

Lemma 4 *Let $(\alpha_1, \dots, \alpha_k)$ and $(\beta_1, \dots, \beta_k)$ be a pair of positive sequences satisfying that*

$$\frac{\alpha_1}{\beta_1} \leq \frac{\alpha_2}{\beta_2} \leq \dots \leq \frac{\alpha_k}{\beta_k}, \quad \text{and} \quad \frac{\beta_1}{\alpha_2} \leq \frac{\beta_2}{\alpha_3} \leq \dots \leq \frac{\beta_{k-1}}{\alpha_k}.$$

Let F be an index set defined by $F \stackrel{\text{def}}{=} \{(1, 1), (2, 2), \dots, (k, k), (2, 1), (3, 2), \dots, (k, k-1)\}$, and $\gamma \in Z^F$ be a vector defined by

$$\begin{aligned} \gamma(1, 1) &= \alpha_1/A, \\ \gamma(i, i) &= (\alpha_1 + \dots + \alpha_i)/A - (\beta_1 + \dots + \beta_{i-1})/B \quad (i = 2, 3, \dots, k), \\ \gamma(i+1, i) &= (\beta_1 + \dots + \beta_i)/B - (\alpha_1 + \dots + \alpha_i)/A \quad (i = 1, 2, \dots, k-1), \end{aligned}$$

where $A = \alpha_1 + \dots + \alpha_k$ and $B = \beta_1 + \dots + \beta_k$. Then the vector γ is non-negative.

Proof.

We can show the lemma in a similar way with the proof of Lemma 2 by removing a_{k+1} . □

Lemma 5 *Assume that the condition*

$$\begin{aligned} (i) \quad & (X^1(1), X^2(1), \dots, X^k(1)) = (Y^1(1), Y^2(1), \dots, Y^k(1)) \text{ and} \\ & (X^2(3), X^3(3), \dots, X^k(3)) = (Y^1(3), Y^2(3), \dots, Y^{k-1}(3)) \end{aligned}$$

is satisfied. Then the following inequalities hold;

$$\begin{aligned} \frac{\psi(X, \{j', j''\})(X^1)}{\psi(Y, \{j', j''\})(Y^1)} &\leq \frac{\psi(X, \{j', j''\})(X^2)}{\psi(Y, \{j', j''\})(Y^2)} \leq \dots \leq \frac{\psi(X, \{j', j''\})(X^k)}{\psi(Y, \{j', j''\})(Y^k)}, \\ \frac{\psi(Y, \{j', j''\})(Y^1)}{\psi(X, \{j', j''\})(X^2)} &\leq \frac{\psi(Y, \{j', j''\})(Y^2)}{\psi(X, \{j', j''\})(X^3)} \leq \dots \leq \frac{\psi(Y, \{j', j''\})(Y^{k-1})}{\psi(X, \{j', j''\})(X^k)}. \end{aligned}$$

Proof. We introduce some notations for simplicity. For any index $\hat{i} \in B_1^n$, we define

$$\begin{aligned} (\eta_{\hat{i}}^1, \eta_{\hat{i}}^2, \dots, \eta_{\hat{i}}^k) &\stackrel{\text{def}}{=} (X^1(\hat{i}; 1), X^2(\hat{i}; 1), \dots, X^k(\hat{i}; 1)), \\ (\bar{\eta}_{\hat{i}}^1, \bar{\eta}_{\hat{i}}^2, \dots, \bar{\eta}_{\hat{i}}^{k+1}) &\stackrel{\text{def}}{=} (X^1(\hat{i}; 3), X^2(\hat{i}; 3), \dots, X^k(\hat{i}; 3), Y^k(\hat{i}; 3)) \\ (\xi_{\hat{i}}^1, \xi_{\hat{i}}^2, \dots, \xi_{\hat{i}}^{k+1}) &\stackrel{\text{def}}{=} (X^1(\hat{i}; 3), X^2(\hat{i}; 3), \dots, X^k(\hat{i}; 3), Y^k(\hat{i}; 3)), \\ (\bar{\xi}_{\hat{i}}^1, \bar{\xi}_{\hat{i}}^2, \dots, \bar{\xi}_{\hat{i}}^k) &\stackrel{\text{def}}{=} (X^1(\hat{i}; 1), X^2(\hat{i}; 1), \dots, X^k(\hat{i}; 1)). \end{aligned}$$

Then, both $(\eta_i^1, \eta_i^2, \dots, \eta_i^k)$ and $(\bar{\eta}_i^1, \bar{\eta}_i^2, \dots, \bar{\eta}_i^{k+1})$ are arithmetic sequences of non-negative integers with common difference -1 . Both of the sequences $(\xi_i^1, \xi_i^2, \dots, \xi_i^{k+1})$ and $(\bar{\xi}_i^1, \bar{\xi}_i^2, \dots, \bar{\xi}_i^k)$ are arithmetic sequences of non-negative integers with common difference 1.

Then, we can show the required result in a similar way with the proof of Lemma 3. □

Above lemmas directly imply the non-negativity of the transition probability of our joint process.

From the above, we have

$$\mathbb{E}[\ell(X', Y')] = \left(1 - \binom{n}{2}^{-1}\right).$$

It implies the following result.

Theorem 2 *The Markov chain \mathcal{M}^2 has the mixing time $\tau_2(\varepsilon)$ satisfying that*

$$\tau_2(\varepsilon) \leq (1/2)n(n-1) \ln(dn/(2\varepsilon)),$$

where d is the average of the values in cells, i.e., $d = N/(2^m n)$.

5 Concluding Remarks

In this paper, we propose two Markov chains for sampling $(m+1)$ -dimensional contingency tables indexed by $\{1, 2\}^m \times \{1, 2, \dots, n\}$. First chain has the uniform distribution as a unique stationary distribution. The stationary distributions of second chain is a conditional multinomial distribution. The mixing times of our chains are bounded by $(1/2)n(n-1) \ln(dn/\varepsilon)$ where d is the average of the values in cells and ε is a given error bound. Thus our chains are rapidly mixing. Our result indicates that the mixing times are independent of the dimension $m+1$ of a contingency table in the case that the size is $2 \times 2 \times \dots \times J$.

References

- [1] A. Agresti, A survey of exact inference for contingency tables, *Statistical Science* 7 (1992) 131–153.
- [2] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, 2002.
- [3] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, in: A. Dold and B. Eckmann (Eds.), *Séminaire de Probabilités XVII 1981/1982*, vol. 986 of Springer-Verlag Lecture Notes in Mathematics, Springer-Verlag, New York, (1983) 243–297.
- [4] S. AOKI AND A. TAKEMURA, Minimal basis for connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals, Technical Report METR 2002-02 Dept. of Mathematical Engineering and Information Physics, Faculty of Engineering, The University of Tokyo, 2002.
- [5] R. Bubley and M. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, 38th Annual Symposium on Foundations of Computer Science IEEE, San Alimitos, (1997) 223–231.
- [6] R. Bubley *Randomized Algorithms : Approximation, Generation, and Counting*, Springer-Verlag, New York, 2001.

- [7] F. R. K. Chung, R. L. Graham, and S. T. Yau, On sampling with Markov chains, *Random Structures and Algorithms* 9 (1996) 55–77.
- [8] P. Diaconis and B. Effron, Testing for independence in a two-way table: new interpretations of the chi-square statistics (with discussion), *Annals of Statistics* 13 (1985) 845–913.
- [9] P. Diaconis and A. Gangolli, Rectangular arrays with fixed margins, in: D. Aldous, P. P. Varaiya, J. Spencer, and J. M. Steele (Eds.), *IMA Volumes on Mathematics and its Applications*, Springer, New York, 72 (1995) 15–41,
- [10] P. Diaconis and L. Saloff-Coste, Random walk on contingency tables with fixed row and column sums, Technical Report, Department of Mathematics, Harvard University, 1995.
- [11] P. Diaconis and B. Strumfels, Algebraic algorithms for sampling from conditional distributions, *The Annals of Statistics* 26 (1998) 363–397.
- [12] M. Dyer and C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables, *Theoretical Computer Sciences* 246 (2000) 265–278.
- [13] M. Dyer, R. Kannan, and J. Mount, Sampling contingency tables, *Random Structures and Algorithms* 10 (1997) 487–506.
- [14] D. Edwards and T. Havranek, A fast procedure for model search in multidimensional contingency table, *Biometrika* 72 (1985) 339–351.
- [15] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyde, Edinburgh, 1934.
- [16] R. A. Fisher, The logic of inductive inference (with discussion), *Journal of Royal Statistical Society* 98 (1935) 39–54.
- [17] D. Hernek, Random generation of $2 \times n$ contingency tables, *Random Structures and Algorithms* 13 (1998) 71–79.
- [18] R. W. Irving and M. R. Jerrum, Three dimensional statistical data security problems, *SIAM Journal on Computing* 23 (1994) 170–184.
- [19] M. R. Jerrum and A. Sinclair, The Markov chain Monte Carlo method: an approach to approximate counting and integration, in *Approximation Algorithm for NP-hard problems*, D. S. Hochbaum (Ed.), PWS publishing, Boston, 1997, 482–520.
- [20] M. Jerrum, A. Sinclair and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries, in *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, ACM press, New York, 2001 712–721.
- [21] R. Kannan, P. Tetali and S. Vempala, Simple Markov chain algorithm for generating bipartite graphs and tournaments, in *8th Annual Symposium on Discrete Algorithms*, ACM-SIAM, San Francisco, California, 1997 193–200.
- [22] C. R. Mehta and N. R. Patel, A network algorithm for performing Fisher’s exact test in $r \times c$ contingency tables, *Journal of the American Statistical Association* 78(1983) 427–434.

- [23] K. Pearson, On the χ^2 test of goodness of fit, *Biometrika* 14 (1922) 186–191.
- [24] W. H. Sewell and A. M. Oewenstein, Community of residencial and occupational choice, *American journal of sociology* 70 (1965) 551–563.
- [25] A. Takemura and S. Aoki, Some characterizations of minimal Markov basis for sampling from discrete conditional distributions Technical Report METR 2002-04, Dept. of Mathematical Engineering and Information Physics, Faculty of Engineering, The University of Tokyo, 2002.