A genuinely polynomial-time algorithm for sampling two-rowed contingency tables (Extended Abstract)

Martin Dyer and Catherine Greenhill
19 March 1998

Abstract

In this paper a Markov chain for contingency tables with two rows is defined. The chain is shown to be rapidly mixing using the path coupling method. The mixing time of the chain is quadratic in the number of columns and linear in the logarithm of the table sum. Two extensions of the new chain are discussed: one for three-rowed contingency tables and one for m-rowed contingency tables. We show that, unfortunately, it is not possible to prove rapid mixing for these chains by simply extending the path coupling approach used in the two-rowed case.

1 Introduction

A *contingency table* is a matrix of nonnegative integers with prescribed positive row and column sums. Contingency tables are used in statistics to store data from sample surveys (see for example [3, Chapter 8]). For a survey of contingency tables and related problems, see [8]. The data is often analysed under the assumption of independence. If the set of contingency tables under consideration is small, this assumption can be tested by applying a chi-squared statistic to each such table (see for example [1, 7]). However, this approach becomes computationally infeasible as the number of contingency tables grows. Suppose that we had a method for sampling almost uniformly from the set of contingency tables with given row and column sums. Then we may proceed by applying the statistic to a sample of contingency tables selected almost uniformly.

The problem of almost uniform sampling can be efficiently solved using the Markov chain Monte Carlo method (see [13]), provided that there exists a Markov chain for the set of contingency tables which converges to the uniform distribution in polynomial time. Here 'polynomial time' means 'in time polynomial in the number of rows, the number of columns and the *logarithm* of the table sum'. If the Markov chain converges in time polynomial in the table sum itself, then we shall say it converges in *pseudopolynomial* time. Approximately counting two-rowed contingency tables is polynomial-time reducible to almost uniform sampling, as can be proved using standard methods. Moreover, the problem of exactly counting the number of contingency tables with fixed row and column sums is known to be $\#P$-complete, even when there are only two rows (see [11]).

The first Markov chain for contingency tables was described in [9] by Diaconis and Saloff-Coste. We shall refer to this chain as the Diaconis chain. For fixed dimensions, they proved that their chain converged in pseudopolynomial time. However, the constants involved grow exponentionally with the number of rows and columns. Some Markov chains for restricted classes of contingency tables have been defined. In [14], Kannan, Tetali and Vempala gave a Markov chain with polynomial-time convergence for the 0-1 case (where every entry in the table is zero or one) with nearly equal margin totals, while Chung, Graham and Yau [6] described a Markov chain for contingency tables which converges in pseudopolynomial time for contingency tables with large enough margin totals. An improvement on this result is the chain described by Dyer, Kannan and Mount [11]. Their chain converges in polynomial time whenever all the row and column sums are sufficiently large, this bound being smaller than that in [6].

In [12], Hernek analysed the Diaconis chain for two-rowed contingency tables using coupling. She showed that this chain converges in time which is quadratic in the number of columns and in the table sum (i.e. pseudopolynomial time). In this paper, a new Markov chain for two-rowed contingency tables is described, and the convergence of the chain is analysed using the path coupling method [4]. We show that the new chain converges to the uniform distribution in time which is quadratic in the number of columns and linear in the *logarithm* of the table sum. Therefore our chain runs in (genuinely) polynomial time, whereas the Diaconis chain does not (and indeed cannot). In the final section we discuss two extensions of the new chain. The first applies to three-rowed contingency tables and the second applies to general contingency tables with m rows. It is not known whether these chains converge rapidly to the uniform distribution, and it is quite possible that they do. However, we show that it is seemingly not possible to prove this simply by extending the path coupling approach used in the two-rowed case.

The structure of the remainder of the paper is as follows. In the next section the path coupling method is reviewed. In Section 3 we introduce notation for contingency tables and describe the Diaconis chain, which converges in pseudopolynomial time. We then outline a procedure which can perform *exact* counting for two-rowed contingency tables in pseudopolynomial time. A new Markov chain for two-rowed contingency tables is described in Section 4 and the mixing time is analysed using path coupling. The new chain is the first which converges in genuinely polynomial time for all two-rowed contingency tables. In Section 5 two extensions of this chain are introduced and discussed.

2 A review of path coupling

In this section we present some necessary notation and review the path coupling method. Let Ω be a finite set and let $\mathcal{M}$ be a Markov chain with state space Ω, transition matrix P and unique stationary distribution π. If the initial state of the Markov chain is x then the distribution of the chain at time t is given by $P_x^t(y) = P^t(x,y)$. The *total variation distance* of the Markov chain from π at time t, with initial state x, is defined by

$$\mathrm{d_{TV}}(P_x^t, \pi) = \frac{1}{2}\sum_{y\in\Omega} |P^t(x,y) - \pi(y)|.$$

A Markov chain is only useful for almost uniform sampling or approximate counting if its total variation distance can be guaranteed to tend to zero relatively quickly, given any initial state. Let $\tau_x(\varepsilon)$ denote the least value T such that $\mathrm{d_{TV}}(P_x^t, \pi) \le \varepsilon$ for all

$t \geq T$. Following Aldous [2], the *mixing time* of $\mathcal{M}$, denoted by $\tau(\varepsilon)$, is defined by $\tau(\varepsilon) = \max\{\tau_x(\varepsilon) : x \in \Omega\}$. A Markov chain will be said to be *rapidly mixing* if the mixing time is bounded above by some polynomial in $\log(|\Omega|)$ and $\log(\varepsilon^{-1})$, where the logarithms are to base e.

There are relatively few methods available to prove that a Markov chain is rapidly mixing. One such method is *coupling*. A *coupling* for $\mathcal{M}$ is a stochastic process (X_t, Y_t) on $\Omega \times \Omega$ such that each of (X_t), (Y_t), considered marginally, is a faithful copy of $\mathcal{M}$. The Coupling Lemma (see for example, Aldous [2]) states that the total variation distance of $\mathcal{M}$ at time t is bounded above by $\text{Prob}[X_t \neq Y_t]$, the probability that the process has not *coupled*. The difficulty in applying this result lies in obtaining an upper bound for this probability. In the *path coupling* method, introduced by Bubley and Dyer [4], one need only define and analyse a coupling on a subset S of $\Omega \times \Omega$. Choosing the set S carefully can considerably simplify the arguments involved in proving rapid mixing of Markov chains by coupling. The path coupling method is described in the next theorem, taken from [10]. Here we use the term *path* to refer to a sequence of elements in the state space, which *need not* form a sequence of possible transitions of the Markov chain.

Theorem 2.1 *Let δ be an integer valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, D\}$. Let S be a subset of $\Omega \times \Omega$ such that for all $(X_t, Y_t) \in \Omega \times \Omega$ there exists a path*

$$X_t = Z_0,\ Z_1, \ldots, Z_r = Y_t$$

between X_t and Y_t where $(Z_l, Z_{l+1}) \in S$ for $0 \leq l < r$ and $\sum_{l=0}^{r-1} \delta(Z_l, Z_{l+1}) = \delta(X_t, Y_t)$. Define a coupling $(X, Y) \mapsto (X', Y')$ of the Markov chain $\mathcal{M}$ on all pairs $(X, Y) \in S$. Suppose that there exists $\beta < 1$ such that

$$\mathbf{E}\left[\delta(X', Y')\right] \leq \beta\, \delta(X, Y)$$

for all $(X, Y) \in S$. Then the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq \frac{\log(D\varepsilon^{-1})}{1 - \beta}.$$

3 Contingency tables

Let $r = (r_1, \ldots, r_m)$ and $s = (s_1, \ldots, s_n)$ be two positive integer partitions of the positive integer N. The set $\Sigma_{r,s}$ of contingency tables with these row and column sums is defined by

$$\Sigma_{r,s} = \left\{ Z \in \mathbb{N}_0^{m \times n} : \sum_{j=1}^{n} Z_{ij} = r_i \text{ for } 1 \leq i \leq m, \sum_{i=1}^{m} Z_{ij} = s_j \text{ for } 1 \leq j \leq n \right\}. \quad (1)$$

The problem of approximately counting the number of contingency tables with given row and column sums is known to be $\#P$-complete even when one of m, n equals 2 (see [11, Theorem 1]). However the 2×2 problem can be solved exactly, as described below.

For 2×2 contingency tables we introduce the notation

$$T_{a,b}^{c} = \Sigma_{(a,c-a),(b,c-b)}$$

where $0 < a,\ b < c$. Now

$$T_{a,b}^{c} = \left\{ \begin{bmatrix} i & (a-i) \\ (b-i) & (c+i-a-b) \end{bmatrix} : \max\{0, a+b-c\} \le i \le \min\{a,b\} \right\}.$$

Hence

$$|T_{a,b}^{c}| = \begin{cases} \min\{a,b\}+1 & \text{if } a+b \le c, \\ c - \max\{a,b\}+1 & \text{if } a+b > c. \end{cases} \tag{2}$$

Choosing an element uniformly at random from $T_{a,b}^{c}$ is accomplished simply by choosing $i \in \{\max\{0, a+b-c\}, \ldots, \min\{a,b\}\}$ uniformly at random and forming the corresponding element of $T_{a,b}^{c}$; that is, the element of $T_{a,b}^{c}$ with i in the north-west corner.

For the remainder of the section, we consider two-rowed contingency tables. Here $m = 2$, and $r = (r_1, r_2)$, $s = (s_1, \ldots, s_n)$ are positive integer partitions of the positive integer N.

We now describe a well-known Markov chain for two-rowed contingency tables. In [9], the following Markov chain for two-rowed contingency tables was introduced. We refer to this chain as the *Diaconis* chain. Let $r = (r_1, r_2)$ and $s = (s_1, \ldots, s_n)$ be two positive integer partitions of the positive integer N. If the current state of the Diaconis chain is $X \in \Sigma_{r,s}$, then the next state X' is obtained using the following procedure. With probability $1/2$ let $X' = X$, otherwise choose two columns uniformly at random, choose $i \in \{1, -1\}$ uniformly at random and add the matrix

$$\begin{bmatrix} i & -i \\ -i & i \end{bmatrix}$$

to the chosen 2×2 submatrix of X. If $X' \notin \Sigma_{r,s}$ then let $X' = X$. It is not difficult to see that this chain is ergodic with uniform stationary distribution (see, for example [12]). This chain was analysed using coupling by Hernek [12]. She proved that the chain is rapidly mixing with mixing rate *quadratic* in the number of columns n and in the table sum N. Hence the Diaconis chain converges in pseudopolynomial time.

To close this section, we show that $|\Sigma_{r,s}|$ can be calculated *exactly* using $O(n\,N^2)$ operations. Hence exact counting is achievable in pseudopolynomial time, and approximate counting is only of value if it can be achieved in polynomial time.

Now $|\Sigma_{r,s}|$ can be calculated using

$$|\Sigma_{r,s}| = \sum_{x} |\Sigma_{(r_1-x, r_2+s_n-x),(s_1,\ldots,s_{n-1})}|, \tag{3}$$

where the sum is over all values of x such that $\max\{0, s_n - r_2\} \le x \le \min\{r_1, s_n\}$. This is a *dynamic programming* problem (see for example, [15]). We can evaluate $|\Sigma_{r,s}|$

exactly using (3), first by solving all the possible 2×2-dimensional problems, then using these results to solve the 2×3-dimensional problems and so on. This procedure costs $O(n\,N^2)$ integer additions, and so $|\Sigma_{r,s}|$ can be calculated exactly in pseudopolynomial time. Moreover, the cost of calculating $|\Sigma_{r,s}|$ in this manner is $O(n)$ lower than the best-known upper bound for the cost of *approximating* $|\Sigma_{r,s}|$ using the Diaconis chain.

4 A new Markov chain for two-rowed contingency tables

For this section assume that $m = 2$. A new Markov chain for two-rowed contingency tables will now be described. First we must introduce some notation. Suppose that $X \in \Sigma_{r,s}$ where $r = (r_1, r_2)$. Given (j_1, j_2) such that $1 \le j_1 < j_2 \le n$ let $T_X(j_1, j_2)$ denote the set $T^c_{a,b}$ where $a = X_{1,j_1} + X_{1,j_2}$, $b = s_{j_1}$ and $c = s_{j_1} + s_{j_2}$. Then $T_X(j_1, j_2)$ is the set of 2×2 contingency tables with the same row and column sums as the 2×2 submatrix of X consisting of the j_1th and j_2th columns of X. (Here the row sums may equal zero.) Let $\mathcal{M}(\Sigma_{r,s})$ denote the Markov chain with state space $\Sigma_{r,s}$ with the following transition procedure. If X_t is the state of the chain $\mathcal{M}(\Sigma_{r,s})$ at time t then the state at time $t+1$ is determined as follows:

(i) choose (j_1, j_2) uniformly at random such that $1 \le j_1 < j_2 \le n$,

(ii) choose $x \in T_X(j_1, j_2)$ uniformly at random and let

$$X_{t+1}(k, j) = \begin{cases} x(k, l) & \text{if } j = j_l \text{ for } l \in \{1, 2\}, \\ X_t(k, j) & \text{otherwise} \end{cases}$$

for $1 \le k \le 2$, $1 \le j \le n$.

Clearly $\mathcal{M}(\Sigma_{r,s})$ is aperiodic. Now $\mathcal{M}(\Sigma_{r,s})$ can perform all the moves of the Diaconis chain, and the Diaconis chain is irreducible (see [12]). Therefore $\mathcal{M}(\Sigma_{r,s})$ is irreducible, so $\mathcal{M}(\Sigma_{r,s})$ is ergodic. Given $X, Y \in \Sigma_{r,s}$ let

$$\phi(X, Y) = \sum_{j=1}^{n} |X_{1,j} - Y_{1,j}|.$$

Then ϕ is a metric on $\Sigma_{r,s}$ which only takes as values the even integers in the range $\{0, \ldots, N\}$. Denote by $\mu(X, Y)$ the minimum number of transitions of $\mathcal{M}(\Sigma_{r,s})$ required to move from initial state X to final state Y. Then

$$0 \le \mu(X, Y) \le \phi(X, Y)/2$$

using moves of the Diaconis chain only (see [12]). However, these bounds are far from tight, as the following shows. Let $K(X, Y)$ be the number of columns which differ in X and Y. The following result gives a bound on $\mu(X, Y)$ in terms of $K(X, Y)$ only.

Lemma 4.1 *If $X, Y \in \Sigma_{r,s}$ and $X \ne Y$ then $\lceil K(X, Y)/2 \rceil \le \mu(X, Y) \le K(X, Y) - 1$.*

Proof. Consider performing a series of transitions of $\mathcal{M}(\Sigma_{r,s})$, starting from initial state X and relabelling the resulting state by X each time, with the aim of decreasing $K(X,Y)$. Each transition of $\mathcal{M}(\Sigma_{r,s})$ can decrease $K(X,Y)$ by at most 2. This proves the lower bound. Now $X \neq Y$ so $K(X,Y) \geq 2$. Let j_1 be the least value of j such that X and Y differ in the jth column. Without loss of generality suppose that $X_{1,j_1} > Y_{1,j_1}$. Then let j_2 be the least value of $j > j_1$ such that $X_{1,j} < Y_{1,j}$. Let $x = \min\{X_{1,j_1} - Y_{1,j_1}, Y_{1,j_2} - X_{1,j_2}\}$. In one move of $\mathcal{M}(\Sigma_{r,s})$ we may decrease X_{1,j_1} and X_{2,j_2} by x and increase X_{1,j_2} and X_{2,j_1} by x. This decreases $K(X,Y)$ by at least 1. The decrease in $K(X,Y)$ is 2 whenever $X_{1,j_1} - Y_{1,j_1} = Y_{1,j_2} - X_{1,j_2}$. This is certainly the case when $K(X,Y) = 2$, proving the upper bound. □

This result shows that the diameter of $\mathcal{M}(\Sigma_{r,s})$ is $(n-1)$, while the diameter of the Diaconis chain is $N/2$. In many cases, N is much larger than n, suggesting that the new chain $\mathcal{M}(\Sigma_{r,s})$ might be considerably more rapidly mixing than the Diaconis chain in these situations. The transition matrix P of $\mathcal{M}(\Sigma_{r,s})$ has entries

$$P(X,Y) = \begin{cases} \sum_{j_1<j_2} \left(\binom{n}{2}|T_X(j_1,j_2)|\right)^{-1} & \text{if } X = Y, \\ \left(\binom{n}{2}|T_X(j_1,j_2)|\right)^{-1} & \text{if } X, Y \text{ differ in } j_1\text{th}, j_2\text{th columns only}, \\ 0 & \text{otherwise.} \end{cases}$$

If all differences between X and Y are contained in the j_1th and j_2th columns only then $T_X(j_1,j_2) = T_Y(j_1,j_2)$. Hence P is symmetric and the stationary distribution of $\mathcal{M}(\Sigma_{r,s})$ is the uniform distribution on $\Sigma_{r,s}$. The Markov chain $\mathcal{M}(\Sigma_{r,s})$ is an example of a *heat bath* Markov chain, as described in [5]. We now prove that $\mathcal{M}(\Sigma_{r,s})$ is rapidly mixing using the path coupling method on the set S of pairs (X,Y) such that $\phi(X,Y) = 2$.

Theorem 4.1 *Let $r = (r_1, r_2)$ and $s = (s_1, \ldots, s_n)$ be two positive integer partitions of the positive integer N. The Markov chain $\mathcal{M}(\Sigma_{r,s})$ is rapidly mixing with mixing time $\tau(\varepsilon)$ satisfying*

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \log(N\varepsilon^{-1}).$$

Proof. Let X and Y be any elements of $\Sigma_{r,s}$. It was shown in [12] that there exists a path

$$X = Z_0, Z_1, \ldots, Z_d = Y \tag{4}$$

such that $\phi(Z_l, Z_{l+1}) = 2$ for $0 \leq l < d$ and $Z_l \in \Sigma_{r,s}$ for $0 \leq l \leq d$, where $d = \phi(X,Y)/2$. Now assume that $\phi(X,Y) = 2$. Without loss of generality

$$Y = X + \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \end{bmatrix}.$$

We must define a coupling $(X,Y) \mapsto (X',Y')$ for $\mathcal{M}(\Sigma_{r,s})$ at (X,Y). Let (j_1,j_2) be chosen uniformly at random such that $1 \leq j_1 < j_2 \leq n$. If $(j_1,j_2) = (1,2)$ or $3 \leq$

$j_1 < j_2 \leq n$ then $T_X(j_1, j_2) = T_Y(j_1, j_2)$. Here we define the coupling as follows: let $x \in T_X(j_1, j_2)$ be chosen uniformly at random and let X' (respectively Y') be obtained from X (respectively Y) by replacing the j_lth column of X (respectively Y) with the lth column of x, for $l = 1, 2$. If $(j_1, j_2) = (1, 2)$ then $\phi(X', Y') = 0$, otherwise $\phi(X', Y') = 2$.

It remains to consider indices (j_1, j_2) where $j_1 \in \{1, 2\}$ and $3 \leq j_2 \leq n$. Without loss of generality suppose that $(j_1, j_2) = (2, 3)$. Let $T_X = T_X(2, 3)$ and let $T_Y = T_Y(2, 3)$. Let $a = X_{1,2} + X_{1,3}$, $b = s_2$ and $c = s_2 + s_3$. Then

$$T_X = T^c_{a,b} \qquad \text{and} \qquad T_Y = T^c_{a+1,b}.$$

Suppose that $a + b \geq c$. Then, relabel the rows of X and Y and swop the labels of the second and third columns of X and Y. Finally interchange the roles of X and Y. Let a', b', c' denote the resulting parameters. Then

$$a' + b' = (c - a - 1) + (c - b) = c - (a + b - c) - 1 < c = c'.$$

Therefore we may assume without loss of generality that $a + b < c$. There are two cases depending on which of a or b is the greater.

Suppose first that $a \geq b$. Then

$$T_X = \left\{ \begin{bmatrix} i_X & (a - i_X) \\ (b - i_X) & (c + i_X - a - b) \end{bmatrix} : 0 \leq i_X \leq b \right\}$$

and

$$T_Y = \left\{ \begin{bmatrix} i_Y & (a + 1 - i_Y) \\ (b - i_Y) & (c + i_Y - a - b - 1) \end{bmatrix} : 0 \leq i_Y \leq b \right\}.$$

Choose $i_X \in \{0, \ldots, b\}$ uniformly at random and let $i_Y = i_X$. Let X' (respectively Y') be obtained from X (respectively Y) by replacing the j_lth column of X (respectively Y) with the lth column of x (respectively y) for $l = 1, 2$. This defines a coupling of $\mathcal{M}(\Sigma_{r,s})$ at (X, Y) for this choice of (j_1, j_2). Here $\phi(X', Y') = 2$.

Suppose next that $a < b$. Then

$$T_X = \left\{ \begin{bmatrix} i_X & (a - i_X) \\ (b - i_X) & (c + i_X - a - b) \end{bmatrix} : 0 \leq i_X \leq a \right\}$$

and

$$T_Y = \left\{ \begin{bmatrix} i_Y & (a + 1 - i_Y) \\ (b - i_Y) & (c + i_Y - a - b - 1) \end{bmatrix} : 0 \leq i_Y \leq a + 1 \right\}.$$

Choose $i_X \in \{0, \ldots, a\}$ uniformly at random and let

$$i_Y = \begin{cases} i_X & \text{with probability } (a - i_X + 1)(a + 2)^{-1}, \\ i_X + 1 & \text{with probability } (i_X + 1)(a + 2)^{-1}. \end{cases}$$

If $i \in \{0, \ldots, a + 1\}$ then

$$\begin{aligned} \text{Prob}\,[i_Y = i] &= \text{Prob}\,[i_X = i] \cdot (a - i + 1)(a + 2)^{-1} \\ &\qquad + \text{Prob}\,[i_X = i - 1] \cdot ((i - 1) + 1)(a + 2)^{-1} \\ &= (a + 1)^{-1} \left((a - i + 1)(a + 2)^{-1} + i(a + 2)^{-1} \right) \\ &= (a + 2)^{-1}. \end{aligned}$$

Therefore each element of $\{0, \ldots, a+2\}$ is equally likely to be chosen, and the coupling is valid. Let x be the element of T_X which corresponds to i_X and let y be the element of T_Y which corresponds to i_Y. Let X', Y' be obtained from X, Y as above. This defines a coupling of $\mathcal{M}(\Sigma_{r,s})$ at (X, Y) for this choice of (j_1, j_2). Again, $\phi(X', Y') = 2$.

Putting this together, it follows that

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

Let $\beta = 1 - \binom{n}{2}^{-1}$. We have shown that $\mathbf{E}\left[\phi(X', Y')\right] = \beta\,\phi(X, Y)$, and clearly $\beta < 1$. Therefore $\mathcal{M}(\Sigma_{r,s})$ is rapidly mixing, by Theorem 2.1. Since $\phi(X, Y) \leq N$ for all $X, Y \in \Sigma_{r,s}$ the mixing time $\tau(\varepsilon)$ satisfies

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2}\log(N\varepsilon^{-1}),$$

as stated. □

5 Two extensions of the two-rowed chain

In this section we describe two natural extensions of the Markov chain $\mathcal{M}(\Sigma_{r,s})$. The first is to three-rowed contingency tables and the second is to general m-rowed contingency tables. We show that the above path coupling approach cannot be successful for either of these chains without major modifications. This does not imply that a different argument might not be found to establish rapid mixing of either chain.

By analogy with the above, we will work with the metric

$$\psi(X, Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} |X_{ij} - Y_{ij}|. \tag{5}$$

Note that, when $m = 2$ we have $\psi(X, Y) = 2\,\phi(X, Y)$ for all $X, Y \in \Sigma_{r,s}$. Also $\psi(X, Y) \geq 4$ whenever $X \neq Y$ and $X, Y \in \Sigma_{r,s}$. Consider the Markov chain which acts on $\Sigma_{r,s}$ by replacing a 2×2 submatrix of the current state, chosen uniformly at random, by another 2×2 matrix with the same row and column sums. This chain is not irreducible: for example, the chain cannot move from X to Y using any sequence of transitions, where

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad Y = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{6}$$

Therefore we must define different Markov chains for contingency tables with more than two rows. We shall use path coupling with respect to the metric ψ. The set S used in the path coupling is defined below.

Definition 5.1 *Let $S \subseteq \Sigma_{r,s} \times \Sigma_{r,s}$ be the set of all pairs $(X, Y) \in \Sigma_{r,s} \times \Sigma_{r,s}$ which satisfy both of the following conditions:*

(i) *there exists* $k \geq 2$ *such that* $\psi(X,Y) = 2k$,

(ii) *there exist indices* $1 \leq i_1, \ldots, i_k \leq m$, $1 \leq j_1, \ldots, j_k \leq n$ *such that the* $\{i_l\}$ *are pairwise distinct, the* $\{j_l\}$ *are pairwise distinct,*

$$Y_{i_l,j_l} = X_{i_l,j_l} + 1 \text{ for } 1 \leq l \leq k, \qquad Y_{i_{l+1},j_l} = X_{i_{l+1},j_l} - 1 \text{ for } 1 \leq l < k,$$

and $Y_{i_1,j_k} = X_{i_1,j_k} - 1$.

5.1 A Markov chain for three-rowed contingency tables

Consider the following natural extension of the Markov chain $\mathcal{M}(\Sigma_{r,s})$ to the case of three-rowed contingency tables. Here $r = (r_1, r_2, r_3)$ and the state space is $\Sigma_{r,s}$. A transition of the extended chain is performed as follows: given a $3 \times n$ matrix X, choose *three* columns of X uniformly at random to give a 3×3 matrix with certain row and column sums. From the set of all 3×3 matrices with these row and column sums, choose a matrix uniformly at random and replace the three selected columns of X with this matrix.

It is not difficult to show that this Markov chain is ergodic and the stationary distribution is the uniform distribution on the set $\Sigma_{r,s}$ of $3 \times n$ contingency tables. Note that selecting an element uniformly at random from a set of 3×3 contingency tables can be easily achieved in polynomial time. Let us investigate the mixing time of this chain using path coupling. We use the metric ψ defined in (5), and the set S of pairs given in Definition 5.1 (where $m = 3$).

We now describe a case which is critical in designing a coupling of the extended chain on elements of S. Let $r = (r_1, r_2, r_3)$ and $s = (s_1, s_2, s_3)$ be two partitions of a positive integer, where the r_i are nonnegative and the s_i are positive. Let $\Sigma_X = \Sigma_{r,s}$ and let $\Sigma_Y = \Sigma_{(r_1+1, r_2-1, r_3), s}$. Suppose that we could define a joint probability distribution $f : \Sigma_X \times \Sigma_Y \to \mathbb{R}$ such that

$$C(f) = \sum_{x \in \Sigma_X} \sum_{y \in \Sigma_Y} f(x,y)\psi(x,y) \leq 2. \tag{7}$$

In order to establish rapid mixing of the extended chain using the path coupling method with the set S and the metric ψ, it suffices to establish (7). Conversely, if we cannot establish (7) then we cannot establish rapid mixing of the chain by path coupling using S and ψ. We do not prove this statement, but instead show that (7) fails.

A joint probability distribution f for $\Sigma_X \times \Sigma_Y$ corresponds to a solution of a related transportation problem, as described in [5]. A joint probability distribution f which minimises $C(f)$ corresponds to an optimal solution Z of the transportation problem. The coupling corresponding to f is called an *optimal coupling*, and $C(f)$ is referred to as the *cost* of an optimal coupling. (Note that an optimal coupling is not the same as a *minimal coupling* referred to in the literature.) The cost of an optimal solution of the related transportation problem is $|\Sigma_X||\Sigma_Y|$ times the cost of an optimal coupling.

We now give a concrete example of such a pair where the cost of the optimal coupling is greater than 2. Let Σ_X be the set of 3×3 contingency tables with row sums $(2, 2, 1)$ and column sums $(2, 2, 1)$, and let Σ_Y be the set of 3×3 contingency tables with row sums $(3,1,1)$ and column sums $(2,2,1)$. Then $|\Sigma_X| = 11$ and $|\Sigma_Y| = 8$. The cost of an optimal solution of the related transportation problem is 180, as can be verified by direct computation. Therefore the cost of an optimal coupling is 45/22, which is greater than 2. This example shows that we cannot easily prove that this chain for three-rowed contingency tables is rapidly mixing using path coupling on the set S with the metric ψ. We cannot conclude from this that the chain is not rapidly mixing or that it might not be possible to prove rapid mixing using another set of pairs S or another metric.

5.2 A Markov chain for m-rowed contingency tables

Here let $r = (r_1, \ldots, r_m)$ and $s = (s_1, \ldots, s_n)$ be two positive integer partitions of the positive integer N. Consider the Markov chain for $\Sigma_{r,s}$ with the following transitions: from current state X, choose two rows of X uniformly at random and replace them by a $2 \times n$ matrix with the same row and column sums, chosen uniformly at random. It is not difficult to show that this chain is ergodic and that the stationary distribution is the uniform distribution on $\Sigma_{r,s}$.

We will investigate the mixing time of this chain using path coupling, using the metric ψ defined in (5) and the set S of pairs given in Definition 5.1. As in the previous subsection, there is one case which is critical in defining a coupling on elements of S. Let $r = (r_1, r_2)$ and $s = (s_1, \ldots, s_n)$ be two partitions of a positive integer, where the r_i are positive and the s_i are nonnegative. Let $\Sigma_X = \Sigma_{r,s}$ and let $\Sigma_Y = \Sigma_{r,s'}$ where $s' = (s_1 + 1, s_2 - 1, s_3, \ldots, s_m)$. Suppose that we could define a joint probability distribution $f : \Sigma_X \times \Sigma_Y \to \mathbb{R}$ such that

$$C(f) = \sum_{x \in \Sigma_X} \sum_{y \in \Sigma_Y} f(x,y)\psi(x,y) \leq 2. \tag{8}$$

If we cannot establish (8) then we cannot prove that the chain is rapidly mixing using path coupling on the set S and the metric ψ. We now demonstrate that (8) fails. As above, a joint probability distribution f for $\Sigma_X \times \Sigma_Y$ corresponds to a solution of the related transportation problem.

We now present a concrete example where the cost of the optimal coupling is greater than 2. Consider the set Σ_X of contingency tables with row sums $(3,2)$ and column sums $(2,2,1)$, and the set Σ_Y of contingency tables with row sums $(3,2)$ and column sums $(3,1,1)$. Then $|\Sigma_X| = 5$ and $|\Sigma_Y| = 4$. It may be verified by direct computation that the cost of an optimal solution of the related transportation problem is 44. Therefore the cost of an optimal coupling is 11/5, which is greater than 2.

This example shows that it is not possible to prove that the Markov chain described in this subsection is rapidly mixing, using the path coupling method on the set S with the metric ψ. It may of course be possible to establish rapid mixing using path coupling with a different set S or a different metric. Note that transitions of this chain may be approximately performed using the Markov chain $\mathcal{M}(\Sigma_{r,s})$ to sample almost uniformly from two-rowed slices of m-rowed contingency tables. This introduces an uncertainty into the transition procedure which results in an increase in the mixing time. Whether this approach is of any practical use depends critically on whether the Markov chain is rapidly mixing, which remains to be seen.

References

[1] J. H. Albert and A. K. Gupta, *Estimation in contingency tables using prior information*, Journal of the Royal Statistical Society, Series B (Methodological), 45 (1983) 1, pp. 60–69.

[2] 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 pp. 243–297.

[3] R. Barlow, *Statistics*, John Wiley and Sons, Chichester, (1989).

[4] R. Bubley and M. Dyer, *Path coupling: A technique for proving rapid mixing in Markov chains*, in *38th Annual Symposium on Foundations of Computer Science*, IEEE, Miami Beach, Florida, 1997 pp. 223–231.

[5] R. Bubley, M. Dyer, and C. Greenhill, *Beating the 2Δ bound for approximately counting colourings: A computer-assisted proof of rapid mixing*, in *9th Annual Symposium on Discrete Algorithms*, ACM–SIAM, San Francisco, California, 1998 pp. 355–363.

[6] F. R. K. Chung, R. L. Graham, and S. T. Yau, *On sampling with Markov chains*, Random Structures and Algorithms, 9 (1996), pp. 55–77.

[7] P. Diaconis and B. Effron, *Testing for independence in a two-way table: new interpretations of the chi-square statistic (with discussion)*, The Annals of Statistics, 13 (1985), pp. 845–913.

[8] P. Diaconis and A. Gangolli, *Rectangular arrays with fixed margins*, in D. Aldous, P. P. Varaiya, J. Spencer, and J. M. Steele, eds., *Discrete Probability and Algorithms*, vol. 72 of *IMA Volumes on Mathematics and its Applications*, Springer-Verlag, New York, 1995 pp. 15–41.

[9] P. Diaconis and L. Saloff-Coste, *Random walk on contingency tables with fixed row and column sums*, Tech. rep., Department of Mathematics, Harvard University, (1995).

[10] M. Dyer and C. Greenhill, *A more rapidly mixing Markov chain for graph colourings*, (1997), (Preprint).

[11] M. Dyer, R. Kannan, and J. Mount, *Sampling contingency tables*, Random Structures and Algorithms, 10 (1997), pp. 487–506.

[12] D. Hernek, *Random generation and counting of rectangular arrays with fixed margins*, (1997), (Preprint).

[13] M. Jerrum and A. Sinclair, *The Markov chain Monte Carlo method: an approach to approximate counting and integration*, in D. Hochbaum, ed., *Approximation Algorithms for NP-Hard Problems*, PWS Publishing, Boston, 1996 pp. 482–520.

[14] R. Kannan, P. Tetali, and S. Vempala, *Simple Markov chain algorithms for generating bipartite graphs and tournaments*, in *8th Annual Symposium on Discrete Algorithms*, ACM–SIAM, San Francisco, California, 1997 pp. 193–200.

[15] A. Kaufmann, *Graphs, Dynamic Programming and finite games*, Academic Press, New York, (1967).