Polynomial-time counting and sampling of two-rowed contingency tables*

Martin Dyer and Catherine Greenhill
School of Computer Studies
University of Leeds
Leeds LS2 9JT
UNITED KINGDOM

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. We prove an upper bound for the mixing time of the chain. The upper bound is quadratic in the number of columns and linear in the logarithm of the table sum. By considering a specific problem, we prove a lower bound for the mixing time of the chain. The lower bound is quadratic in the number of columns and linear in the logarithm of the number of columns. A fully polynomial randomised approximation scheme for this problem is described.

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, 20]). 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 [16]), 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

*An earlier version of this paper appeared in the 25th Annual EATCS International Colloquium on Automata, Languages and Programming, July 1998

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 we prove in Section 3. 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 [13]).

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 converges 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 [17], 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 [13]. 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 [15], 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). By considering a specific example, we prove a lower bound for the new Markov chain which is quadratic in the number of columns and linear in the logarithm of the number of columns.

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 show that approximate counting of two-rowed contingency tables is polynomial-time reducible to almost uniform sampling. We describe the Diaconis chain, which converges in pseudopolynomial time. Although this chain can be used for approximate counting, we present 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. A lower bound for the mixing time of this chain is developed in Section 5.

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 $\mathrm{d_{TV}}(P_x^t, \pi) \leq \varepsilon$ 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 $\mathrm{Prob}[X_t \neq Y_t]$, 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 [12]. 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 [13, 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

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

Hence

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

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 show that approximately counting two-rowed contingency tables with given row and column sums is polynomial-time reducible to almost uniform sampling. First let us make this statement more precise. Let ε, δ be such that $0 < \varepsilon, \delta < 1$. A *fully polynomial randomised approximation scheme* (or FPRAS) [18] for $|\Sigma_{r,s}|$ is a randomised algorithm which runs in time polynomial in n, $\log(N)$, ε^{-1} and $\log(\delta^{-1})$ to produce an estimate Z for $|\Sigma_{r,s}|$, such that

$$\text{Prob}\left[(1-\varepsilon)|\Sigma_{r,s}| \leq Z \leq (1+\varepsilon)|\Sigma_{r,s}|\right] \geq 1 - \delta.$$

Suppose that $\mathcal{M}$ is a rapidly mixing Markov chain for two-rowed contingency tables, and let $\tau(\varepsilon)$ denote the mixing time of $\mathcal{M}$. Then $\tau(\varepsilon)$ is bounded above by a polynomial in n, $\log(N)$ and $\log(\varepsilon^{-1})$. We now describe an FPRAS for $\Sigma_{r,s}$ which uses the Markov chain $\mathcal{M}$ to perform almost uniform sampling from $\Sigma_{r,s}$. This FPRAS differs from most of the known schemes for other combinatorial problems, as its complexity depends upon *number size* (that is, the column sums themselves and not just the number of columns). For this reason, we present a brief description. Some proof details are omitted, as the arguments involved are standard [10, 16].

Let R be defined by $R = \sum_{q=3}^{n} \lceil \log_2(s_q) \rceil$. Then $R < n \log(N)$. We estimate $|\Sigma_{r,s}|$ in R steps, and only describe the first step in detail. Let $\mathcal{U}_j$ be defined by

$$\mathcal{U}_j = \{X \in \Sigma_{r,s} : X_{jn} \geq \lceil s_n/2 \rceil\}$$

for $j = 1, 2$. Let M be defined by

$$M = \lceil 150 e^2 R^2 \varepsilon^{-2} \log(3R\delta^{-1}) \rceil.$$

Using the Markov chain $\mathcal{M}$, produce a sample $S \subseteq \Sigma_{r,s}$ of size M as follows:

$S := \emptyset$;
let x be some arbitrarily chosen element of $\Sigma_{r,s}$;
let $T := \tau(\varepsilon/(15Re^2))$;
for $j := 1$ to M do
    perform T steps of the Markov chain $\mathcal{M}$, starting from initial state x;
    let $S := S \cup \{X\}$, where X is the final state of the chain;
enddo;

Define $J \in \{1, 2\}$ by

$$J = \begin{cases} 1 & \text{if } |S \cap \mathcal{U}_1| \geq M/2, \\ 2 & \text{otherwise.} \end{cases}$$

Let $Z_1 = |S \cap \mathcal{U}_J| M^{-1}$. Then $Z_1 \geq 1/2$. We take Z_1 as our estimate for ρ_1, where $\rho_1 = |\mathcal{U}_J||\Sigma_{r,s}|^{-1}$. Let r' be obtained from r by subtracting $\lceil s_n/2 \rceil$ from r_J. Finally, let s' be defined by

$$s' = \begin{cases} (s_1, \ldots, s_{n-1}, \lfloor s_n/2 \rfloor) & \text{if } s_n > 1, \\ (s_1, \ldots, s_{n-1}) & \text{if } s_n = 1 \end{cases}.$$

Then $|\mathcal{U}_J| = |\Sigma_{r',s'}|$, so this quantity can be estimated using the same method. The process terminates after R steps, where it remains to calculate the number of 2×2 contingency tables with given row and column sums. Let σ be this number, which can be evaluated exactly. Let Z_i be the estimate for ρ_i which we obtain at the ith step. Then $Z = \sigma (Z_1 \cdots Z_R)^{-1}$ is our estimate for $|\Sigma_{r,s}|$, while $|\Sigma_{r,s}| = \sigma (\rho_1 \cdots \rho_R)^{-1}$ by construction.

We now outline the steps involved in proving that this procedure is indeed an FPRAS. Let $\hat{\rho}_i = \mathbf{E}[Z_i]$ for $1 \leq i \leq R$. By choice of M and the simulation length of the Markov chain, we can prove the following:

(i) for $1 \leq i \leq R$, $|\rho_i - \hat{\rho}_i| \leq \varepsilon/(15Re^2)$,

(ii) for $1 \leq i \leq R$, $\text{Prob}[\hat{\rho}_i \geq 1/(2e^2)] \geq 1 - \delta/(3R)$,

(iii) if $\hat{\rho}_i \geq 1/(2e^2)$ for some i such that $1 \leq i \leq R$, then $|\rho_i - \hat{\rho}_i| \leq \varepsilon/(5R)\hat{\rho}_i$,

(iv) if $\hat{\rho}_i \geq 1/(2e^2)$ for some i such that $1 \leq i \leq R$, then

$$\text{Prob}[|Z_i - \hat{\rho}_i| > \varepsilon/(5R)\hat{\rho}_i] \leq 2\delta/(3R),$$

(v) with probability at least $1-\delta$, we have

$$|(Z_1 \dots Z_R)^{-1} - (\rho_1 \dots \rho_R)^{-1}| \leq \varepsilon(\rho_1 \dots \rho_R)^{-1}.$$

It is now not difficult to see that

$$\text{Prob}\left[(1-\varepsilon)|\Sigma_{r,s}| \leq Z \leq (1+\varepsilon)|\Sigma_{r,s}|\right] \geq 1-\delta.$$

(The structure of this argument is standard, see [10, 16].) Thus, with high enough probability, we have estimated $|\Sigma_{r,s}|$ to within the required accuracy.

In order to estimate the complexity of this procedure, assume that the mixing time of the Markov chain used in the ith step of the procedure is bounded above by $\tau(\varepsilon)$, the mixing time of the Markov chain used in the first step. This is reasonable since the number of columns is non-increasing as the procedure progresses. The total number of Markov chain simulations used in this procedure is bounded above by

$$R\,M\,\tau(\varepsilon/(15Re^2)).$$

Since $\mathcal{M}$ is rapidly mixing, and by definition of M and R, this expression is polynomial in n, $\log(N)$, ε^{-1} and $\log(\delta^{-1})$. This proves the existence of an FPRAS for approximately counting contingency tables. In other words, approximate counting of two-rowed contingency tables is polynomial-time reducible to almost uniform sampling.

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 $s = (s_1, \dots, s_n)$ 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 $X' \in \Sigma_{r,s}$ is produced using the following procedure: with probability $1/2$, let $X' = X$. Otherwise, choose choose (j_1, j_2) uniformly at random such that $1 \leq j_1 < j_2 \leq n$, and choose $i \in \{1, -1\}$ uniformly at random. Form X' from X by adding the matrix

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

to the 2×2 submatrix of X consisting of the j_1th and j_2th columns 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 [15]). This chain was analysed using coupling by Hernek [15]. 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 how to calculate $|\Sigma_{r,s}|$ *exactly* using $O(n\,N)$ operations. The method, due to Diane Hernek [14], is an efficient dynamic programming approach based on a certain recurrence. This shows that exact counting is achievable in pseudopolynomial time, and approximate counting is only of value if it can be achieved in polynomial time. The new chain described in the next section is the first Markov chain for general two-rowed contingency tables which provably converges in polynomial time.

Let $R = \min\{r_1, r_2\}$, and for $1 \le i \le R$, $1 \le j \le n$ let

$$\mathcal{S}_{ij} = \left\{ (x_1, \ldots, x_j) : \sum_{k=1}^{j} x_k = i \text{ and } 0 \le x_k \le s_k \text{ for } 1 \le k \le j \right\}.$$

Let $A_{ij} = |\mathcal{S}_{ij}|$ for $0 \le i \le R$, $0 \le j \le n$. Then $|\Sigma_{r,s}| = A_{Rn}$. Let $A_{0j} = 1$ for $0 \le j \le n$, and let $A_{i0} = 0$ for $1 \le i \le R$. It is not difficult to show that the following recurrence holds:

$$A_{ij} = \begin{cases} A_{i,j-1} + A_{i-1,j} & \text{if } i-1 < s_j, \\ A_{i,j-1} + A_{i-1,j} - A_{i-s_j-1,j-1} & \text{if } i-1 \ge s_j. \end{cases}$$

This leads to the following $O(nN)$-time algorithm for calculating $|\Sigma_{r,s}|$.

```
Begin
  for j := 0 to n do
    A_{0j} := 1;
  endfor;
  for i := 1 to R do
    A_{i0} := 0;
  endfor;
  for i := 1 to R do
    for j := 1 to n do
      A_{ij} := A_{i,j-1} + A_{i-1,j};
      if i - 1 ≥ s_j then
        A_{ij} := A_{ij} - A_{i-s_j-1,j-1};
      endif;
    endfor;
  endfor;
  return A_{Rn};
End.
```

As mentioned above, this procedure has running time $O(n\,N)$. Thus it saves a factor of nN compared with the best known upper bound of $O(n^2\,N^2)$ for the cost of generating a single sample from $\Sigma_{r,s}$ 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 [15]). 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

$$0 \le \mu(X,Y) \le \phi(X,Y)/2 \tag{3}$$

using moves of the Diaconis chain only (see [15]). 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 $X \neq Y$ 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 $K(X,Y) \ge 2$. 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 at most $(n-1)$. Now (3) implies that the diameter of the Diaconis chain is at most $\lfloor N/2 \rfloor$. By considering the set of 2×2 contingency tables with row and column sums given by $(\lfloor N/2 \rfloor, \lceil N/2 \rceil)$, we see that $\lfloor N/2 \rfloor$ is also a lower bound for the diameter of the Diaconis chain; that is, the diameter of the Diaconis chain is *exactly* $\lfloor N/2 \rfloor$. 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

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

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 [15] 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

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

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 $X, Y \in \Sigma_{r,w}$ the mixing time $\tau(\varepsilon)$ satisfies

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

as stated. □

Attempts have been made to extend this Markov chain to act on general m-rowed contingency tables, so far without success. The problem seems much harder, even when restricted to three-rowed contingency tables. See [11] for more details.

5 A lower bound for the mixing time

In this section we find a lower bound for the mixing time of the Markov chain $\mathcal{M}$ for two-rowed contingency tables. We proceed by considering a specific sequence of contingency tables, defined below. In the chosen example we have $N = 2(n-1)$, and we prove the lower bound

$$\frac{n(n-1)}{6} \log\left(\frac{n-1}{8}\right) \leq \tau(e^{-1}).$$

Taking the upper and lower bounds together shows that

$$\tau(e^{-1}) = \Theta(n^2 \log(n))$$

for this example. Of course we do not always have $N = \Theta(n)$. It is not known whether, in general, the upper bound $O(n^2 \log(N))$ is tight, or whether $\Omega(n^2 \log(n))$ is the correct lower bound.

We now define the set of contingency tables to be considered. It is the set $\Sigma_{r,s}$ of contingency tables with row sums $r = (n-1, n-1)$ and column sums $s = (n-1, 1, \ldots, 1)$. Suppose that $X \in \Sigma_{r,s}$. Then $0 \leq X_{11} \leq n-1$. Moreover, there are exactly $\binom{n-1}{i}$ elements $X \in \Sigma_{r,s}$ such that $X_{11} = i$. It follows easily that $|\Sigma_{r,s}| = 2^{n-1}$. The distribution of the top-left element X_{11} is *binomial* when X is selected uniformly at random from $\Sigma_{r,s}$.

We analyse the sequence $\{X_k\}$ obtained by simulating the Markov chain $\mathcal{M}(\Sigma_{r,s})$, starting from the initial state X_0 given by

$$X_0 = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ n-1 & 0 & \cdots & 0 \end{bmatrix}. \tag{5}$$

We will focus particularly on the top-left element of each member of this sequence. For this reason, let $\{Y_k\}$ be the sequence defined by $Y_k = (X_k)_{11}$ for $k \geq 0$. Then $0 \leq Y_k \leq n-1$ for $k \geq 0$. Informally, the distribution of X_k cannot be close to uniform until the distribution of Y_k is close to binomial.

Let

$$\beta = 1 - \frac{2}{n(n-1)}.$$

Using standard methods, it is possible to prove that

$$\mathbf{E}[Y_k] = \frac{n-1}{2}\left(1 - \beta^k\right) \tag{6}$$

for $k \geq 0$. Hence $\{\mathbf{E}[Y_k]\}$ is an increasing sequence with limit $(n-1)/2$, the expected value in the binomial distribution. Again using standard methods, one can prove that

$$\operatorname{var}(Y_k) = \frac{n-1}{4}\left(1 - (n-1)\beta^{2k} + (n-2)(2\beta - 1)^k\right) \tag{7}$$

for $k \geq 0$. In particular, $\text{var}(Y_k) \leq (n-1)/4$ for all $k \geq 0$. Using these results, we can prove the following lower bound on the mixing time.

Theorem 5.1 *Let $\tau(\varepsilon)$ denote the mixing time of the Markov chain $\mathcal{M}(\Sigma_{r,s})$. Then*

$$\frac{n(n-1)}{6} \log\left(\frac{n-1}{8}\right) \leq \tau(e^{-1}).$$

Proof. Recall the set $\Sigma_{r,s}$ defined at the start of this section, and the initial state X_0, defined in (5). Let P^k denote the distribution of the Markov chain $\mathcal{M}(\Sigma_{r,s})$, after performing k steps from initial state X_0. Denote by π the uniform distribution on $\Sigma_{r,s}$, and let $d_k = \mathrm{d_{TV}}(P^k, \pi)$, the total variation distance between π and P^k. We bound d_k from below by comparing the probability that the top-left element of X is at least $(n-1)/2$, when X is chosen according to π and P^k respectively. Formally, let S be defined by

$$S = \{X \in \Sigma_{r,s} : X_{11} \geq (n-1)/2\}.$$

Then $|\pi(S) - P^k(S)| \leq d_k$, by definition of total variation distance. Clearly $\pi(S) \geq 1/2$. Using Chebyshev's bound and standard arguments, it is possible to show that

$$P^k(S) \leq \frac{1}{(n-1)\beta^{2k}},$$

where $\beta = 1 - 2/(n(n-1))$. Therefore

$$\frac{1}{2} - \frac{1}{(n-1)\beta^{2k}} \leq |\pi(S) - P^k(S)| \leq d_k.$$

Suppose that $k = \tau(e^{-1})$, where $\tau(\varepsilon)$ is the mixing time of the Markov chain $\Sigma_{r,s}$. Then $d_k \leq e^{-1}$, which implies that

$$\frac{1}{2} - \frac{1}{(n-1)\beta^{2k}} \leq e^{-1}.$$

After some rearranging, this implies that

$$k \geq \frac{\beta \log\left(\frac{n-1}{8}\right)}{2(1-\beta)} \geq \frac{n(n-1)}{6} \log\left(\frac{n-1}{8}\right).$$

Here the final inequality uses the fact that $\beta \geq 2/3$. This proves the theorem. □

Taking the lower bound for $\tau(e^{-1})$ proved in Theorem 5.1, together with the upper bound for the mixing time proved in 4.1, we see that $\tau(e^{-1}) = \Theta(n^2 \log(n))$. In the chosen example, the table sum N satisfies $N = \Theta(n)$. Therefore, this lower bound does not tell us whether the $\log(n)$ term should be $\log(N)$ in general, or whether $\Omega(n^2 \log(n))$ is the true lower bound. We conjecture that $\Theta(n^2 \log(n))$ is true in general, for the following reason. We feel that if every 2×2 submatrix has been visited some *constant* number of times, then this should ensure that the resulting table is very close to random. This requires $\Theta(n^2 \log(n))$ steps, by results for the coupon-collectors problem (see, for example [19, pp.57–63]).

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, San Alimitos, 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, New York–Philadelphia, 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 A. Frieze, *Computing the volume of convex bodies: A case where randomness provably helps*, Proceedings of Symposia in Applied Mathematics, 44 (1991), pp. 123–169.

[11] M. Dyer and C. Greenhill, *A genuinely polynomial time algorithm for sampling two-rowed contingency tables*, in *25th International Colloquium on Automata, Languages and Programming*, Aalborg, Denmark, 1998 pp. 339–350.

[12] M. Dyer and C. Greenhill, *A more rapidly mixing Markov chain for graph colourings*, Random Structures and Algorithms, 13 (1998), pp. 285–317.

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

[14] D. Hernek, Private Communication, (1998).

[15] D. Hernek, *Random generation of* $2 \times n$ *contingency tables*, Random Structures and Algorithms, 13 (1998), pp. 71–79.

[16] 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.

[17] 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, New York–Philadelphia, 1997 pp. 193–200.

[18] R. M. Karp and M. Luby, *Monte-Carlo algorithms for enumeration and reliability problems*, in *24th Annual Symposium on Foundations of Computer Science*, IEEE, San Alimitos, 1983 pp. 56–64.

[19] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, (1995).

[20] J. N. K. Rao and A. J. Scott, *The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables*, Journal of the American Statistical Association, 76 (1981) 374, pp. 221–230.