Sampling contingency tables

Koko K. Kayibi[a], S. Pirzada[b,*], T.A. Chishti[b]

[a] *Department of Mathematics and Physics, University of Hull, United Kingdom*
[b] *Department of Mathematics, University of Kashmir, Srinagar, India*

Abstract

Let Ω be the set of all $m \times n$ matrices, where r_i and c_j are the sums of entries in row i and column j, respectively. Sampling efficiently uniformly at random elements of Ω is a problem with interesting applications in Combinatorics and Statistics. To calibrate the statistic χ^2 for testing independence, Diaconis and Gangolli proposed a Markov chain on Ω that samples uniformly at random contingency tables of fixed row and column sums. Although the scheme works well for practical purposes, no formal proof is available on its rate of convergence. By using a canonical path argument, we prove that this Markov chain is fast mixing and the mixing time $\tau_x(\epsilon)$ is given by

$$\tau_x(\epsilon) \leq 2c_{max}(mn)^4 \ln(c_{max}\epsilon^{-1}),$$

where $c_{max} - 1$ is the maximal value in a cell.

1. Introduction

An $m \times n$ *contingency table* is an $m \times n$ array where the sum of entries in row i is r_i and the sum of entries in column j is c_j. Let the *size* of a contingency table be the multiset of its rows and columns sums r_i and c_j. The combinatorial problems involving contingency tables are varied and interesting. They include counting magic squares and enumerating permutations by descent patterns and counting double cosets of symmetric groups. Contingency tables are also useful in the study of induced representations, tensor products and symmetric functions. See [1] for a comprehensive introduction on these topics. Moreover, random sampling contingency tables is a problem that has statistical applications, as illustrated in [2,3]. For graph theoretic concepts, we refer to [4].

While entries in contingency tables may be any non-negative integers, the case where entries are 0 or 1 had deserved a special attention in the literature [1–3,5–12]. Indeed, (0, 1) matrices are much researched upon for its relevance in various fields of sciences, such as network modeling in ecology, social sciences, chemical compounds and biochemical networks in the cell, where they serve to model bipartite graphs with fixed degree sequences. Although random sampling contingency tables and (0, 1) matrices problems seem similar, the present paper focuses exclusively on the former.

Let Ω be the set of $m \times n$ contingency tables. One way to sample a table at random is to construct a random walk that starts at a given element of Ω, then moves to another element according to some simple rule which changes one element to another. After i steps, one outputs the element reached by the random walk and considers this element, the ith state reached, as a random sample. Such a random walk is a Markov chain.

Although this description seems straightforward, there are two issues at stake. First of all, the rule of transition describing the *local perturbation* from one element to another has to be simple enough so that it can be implemented. Moreover, the local perturbation must be such that, starting from any element, any other element has a chance to be visited. If the Markov chain is viewed as a random walk on a digraph $\mathcal{G}$ so that the elements of Ω are vertices of $\mathcal{G}$, the last condition is equivalent to require that $\mathcal{G}$ be strongly connected. In [1,2], the following Markov chain is proposed. Take an $m \times n$ contingency table T where entries in row i add up to r_i and entries in column j add up to c_j. Choose at random two rows i and i' and two columns j and j'. Then make the following local perturbation on table T, changing it into the array T' by the rule $T'_{ij} = T_{ij} + 1$, $T'_{i'j} = T_{i'j} - 1$, $T'_{ij'} = T_{ij'} - 1$, and $T'_{i'j'} = T_{ij} + 1$. Move from T to T' if T' has no negative entry, and stay on T otherwise. It can be easily proved that this operation is a local perturbation that defines a random walk which visits all the contingency tables having row sums and column sums r_i and c_j, respectively.

There is another issue to be addressed which is as follows. Since the successive states of a Markov chain are not independent, an unbiased sample can only be obtained if the chain reaches stationarity, the stage when the probability of sampling a particular configuration is fixed in time. The *mixing time* of a Markov chain is the number of steps required to reach that stationary state up to a defined ϵ. The information about the mixing time of a particular Markov chain is important to avoid to get a biased sample (if stationarity is not reached), or to avoid the computational cost of running the chain more than necessary. The present paper solves this problem by showing that the Markov chain proposed in [1,2,13] mixes fast. That is, the stationary state is reached after a number of steps which is polynomial on mn, the size of the contingency table having fixed row and column sums r_i and c_j, respectively. For the closely related problem of sampling (0, 1) matrices, some partial results are available. Indeed, while the mixing time of the general problem is still an open problem, it had been proven by Erdos et al. [14] to be fast mixing for semi-regular degree sequences only.

2. Basics on Markov chains

Let $\mathcal{M}$ be a Markov chain on a set of states Ω. Let P be the matrix of transitions from one state to another. One may visualize the Markov chain $\mathcal{M}$ as a weighted directed graph $\mathcal{G}$, where the states are the vertices of $\mathcal{G}$ and there is an edge of weight $P(x, y)$ if the transition from state x to state y has probability $P(x, y)$. A Markov chain is *irreducible* if for all pairs of states x and y, there is an integer t, depending on the pair (x, y), such that $P^t(x, y) > 0$. In terms of the digraph $\mathcal{G}$, the Markov chain is irreducible if there is a directed path between every pair of vertices of $\mathcal{G}$. A Markov chain is *aperiodic* if for all states x, there is an integer t such that for all $t' \geq t$, $P^{t'}(x, x) > 0$. That is, after sufficient number of iterations, the chain has a positive probability to stay on x at every subsequent step. This ensures that the return to state x is not periodic. In terms of the graph $\mathcal{G}$, this can be achieved by having a loop at every vertex. A Markov chain is *ergodic* if it is irreducible and aperiodic. If P is the matrix of transitions of an ergodic Markov chain, it can be shown that there is an integer t such that for all the pairs of states (x, y), $P^t(x, y) > 0$. (Notice that t does not depend on the pair). Also, it can be proven that for every ergodic Markov chain with transition matrix P, the largest eigenvalue of P only occurs once and is equal to 1 (Perron–Frobenius Theorem). Using this, it can be proved that there is a unique probability vector π, such that $\pi P = \pi$. The vector π is the *stationary* distribution of $\mathcal{M}$. A chain is *reversible* if $P(x, y)\pi(x) = P(y, x)\pi(y)$ for every pair of states x and y. If a chain is reversible and the matrix P is symmetric, then $\pi(x) = \pi(y)$ for every pair x and y. The stationary state is then said to be *uniform*. We say a matrix P is *irreducible, ergodic, aperiodic, etc.* if the Markov chain ruled by P is irreducible, ergodic, aperiodic, etc.

Let $\mathcal{M}$ be an ergodic Markov chain defined on a finite set of states Ω, with a transition matrix P and stationary distribution π. Starting the chain from an initial state x, we would like to measure $\Delta_x(t)$, the distance between the

distribution at time t and the stationary distribution. More formally, if $P^t(y|x)$ represents the probability that, at time t, the chain is at state y given initial state x, and $\pi(y)$ represents the probability that the chain is at state y at stationarity, then the *variation distance*, denoted by $\Delta_x(t)$, is defined as

$$\Delta_x(t) = \frac{1}{2} \sum_{y \in \Omega} |P^t(y|x) - \pi(y)|.$$

A converging chain is one such that $\Delta_x(t) \to 0$ as $t \to \infty$ for all initial states x. The *rate of convergence* is measured by $\tau_x(\epsilon)$, the time required to reduce the variation distance to ϵ given an initial state x.

$$\tau_x(\epsilon) = min\{t : \Delta_x(t) \leq \epsilon, \quad for \ all \ t' \geq t\}.$$

The *mixing time* of the chain, denoted by $\tau(\epsilon)$, is defined as $max_{x \in \Omega}$, the maximum being over all the initial points x. A Markov chain is said to be *rapidly mixing* if its mixing time is bounded above by a polynomial in the size of the input and $\frac{1}{\epsilon}$.

Let $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of the transition matrix P. The *spectral gap* of the matrix P is defined as $max\{1 - \lambda_2, 1 - |\lambda_n|\}$. If $P(x, x) < \frac{1}{2}$ for all x, then $\lambda_n > 0$, and therefore smaller than λ_2. The spectral gap of P is then the real number $\lambda_1 - \lambda_2$, that is, $1 - \lambda_2$. It can be shown that largest the gap, the faster is the mixing time of the chain. The analysis of the mixing time of a Markov chain is based on the intuition that a random walk on the graph $\mathcal{G}$ mixes fast (i.e., reaches all the states quickly) if $\mathcal{G}$ has no bottleneck. That is, there are no cuts between any set of vertices S to its complement, which blocks the flow of the Markov chain and thus prevents the Markov chain from reaching easily some states. See [10,12,15–17] for a better exposition on the topic. To make this more formal, we need some preliminary definitions which conforms with [12].

Denoting the probability of x at stationarity by $\pi(x)$ and the probability of moving from x to y by $P(x, y)$, the *capacity* of the arc $e = (x, y)$, denoted by $c(e)$, is given by

$$c(e) = \pi(x)P(x, y).$$

Let $\mathcal{P}_{x,y}$ denote the set of all simple paths p from x to y (paths that contain every vertex at most once). A *flow* in $\mathcal{G}$ is a function ϕ, from the set of simple paths to the reals, such that

$$\sum_{p \in \mathcal{P}_{x,y}} \phi(p) = \pi(x)\pi(y),$$

for all vertices x, y of $\mathcal{G}$ with $x \neq y$. An *arc-flow* along an arc e, denoted by $\phi'(e)$, is then defined as

$$\phi'(e) = \sum_{p \ni e} \phi(p).$$

For a flow ϕ, a measure of existence of an overload along an arc is given by the quantity $\rho(e)$, where

$$\rho(e) = \frac{\phi'(e)}{c(e)},$$

and the *cost* of the flow ϕ, denoted by $\rho(\phi)$, is given by

$$\rho(\phi) = \max_e \rho(e).$$

If a network $\mathcal{G}$ representing a Markov chain can support a flow of low cost, then it cannot have any bottlenecks, and hence its mixing time should be small. This intuition is confirmed by the following theorem [12].

Theorem 1 (*[12]*). *Let $\mathcal{M}$ be an ergodic reversible Markov chain with holding probabilities $P(x, x) \geq \frac{1}{2}$ at all states x. The mixing time of $\mathcal{M}$ satisfies*

$$\tau_x(\epsilon) \leq \rho(\phi)|p|\left(ln\frac{1}{\pi(x)} + ln\frac{1}{\epsilon}\right),$$

where $|p|$ is the length of a longest path carrying non-zero flow in ϕ.

Thus, one way to prove that a Markov chain mixes fast is to produce a flow along some paths, where the paths and the maximal overload on edges are polynomials on the size of the problem.
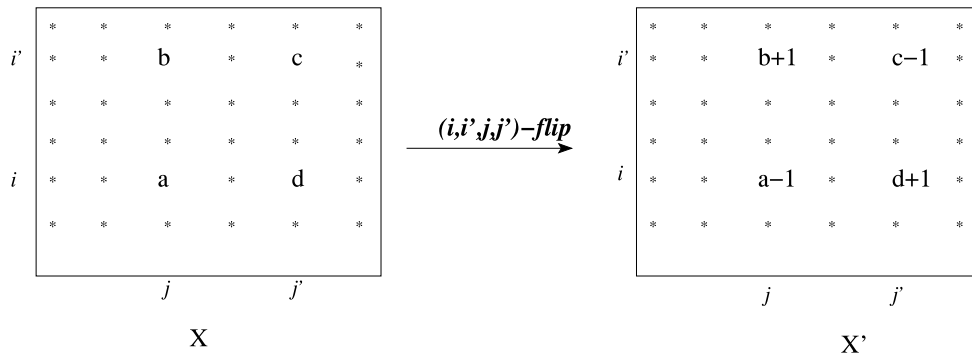
Fig. 1. A flip on the site $[i, i', j, j']$.

3. Markov chains on the set of contingency tables

Diaconis et al. [1,2] define local perturbations that transform a contingency table into another of the same row and column sums. We now describe it in more detail while redefining some terms for completeness. Let X be an $m \times n$ contingency table of row and column sums r_i and c_j, respectively. Let X_{ij} be the entry on the ith row and j^{th} column. That is, X_{ij} is the entry in the cell (i, j). We totally order the cells of X lexicographically. That is, $(i, j) > (k, l)$ if $i > k$. If $i = k$, then $(i, j) > (k, l)$ with $j > l$. A *site*, denoted by $[i, i', j, j']$, is the set of four cells $\{(i, j), (i', j), (i, j'), (i', j')\}$, with $i \neq i'$ and $j \neq j'$. A *flip* on $[i, i', j, j']$ is a transformation that changes the contingency table X into X' by the operations

$$X'_{i,j} = X_{i,j} + 1$$
$$X'_{i',j} = X_{i',j} - 1$$
$$X'_{i',j'} = X_{i',j'} + 1$$
$$X'_{i,j'} = X_{i,j'} - 1,$$

while preserving all the other entries. See Fig. 1 for an illustration.

The following result given in [1,2], shows that the Markov chain on Ω is irreducible.

Theorem 2 ([1,2]). *Given two contingency tables X and Y in Ω, there is a sequence of flips which transforms X to Y.*

Let $\mathcal{G}$ denote the graph whose vertices are all $m \times n$ contingency tables of row and column sums r_i and c_j, respectively, and there is an edge from a vertex X to Y if and only if there is a site $[i, i', j, j']$ such that a flip on $[i, i', j, j']$ changes X to Y. If X is any vertex of $\mathcal{G}$, it is routine to check that X has at most $(mn)^2$ adjacent neighbors. Indeed, the vertex Y is an adjacent neighbor of X if there is a flip changing X to Y. Now, given a cell (i, j) on the contingency table X, there are at most $(m - 1)(n - 1)$ sites involving (i, j). Considering all mn points on the array X gives a total of at most $(mn)^2$ possible sites.

For the sake of convenience, in what follows, we use X or x interchangeably. We write X when we need to stress that we are talking about a contingency table and we write x if we need to stress that we are talking about the vertex that represents X in the graph $\mathcal{G}$. Now, we formally define a random walk on $\mathcal{G}$ as follows. Start at any vertex z. If the walk is at x, choose at random an adjacent neighbor y of x, and move from x to y with probability $P_{x,y} = \frac{1}{2(mn)^2}$, or stay at x otherwise. If P denotes the transition matrix and $P(x)$ is the row of P corresponding to the state x, it is routine to check that the entries of $P(x)$ add up to 1. That is, $P(x)$ is a probability vector. Moreover, by Theorem 2, the matrix P is irreducible. Also, since there is a loop on every point x, the matrix P is aperiodic. Hence the random walk is an ergodic Markov chain. Finally, P is a symmetric matrix. Thus, the Markov chain has a stationary uniform distribution π. We call this Markov chain as the *Diaconis Markov chain*.
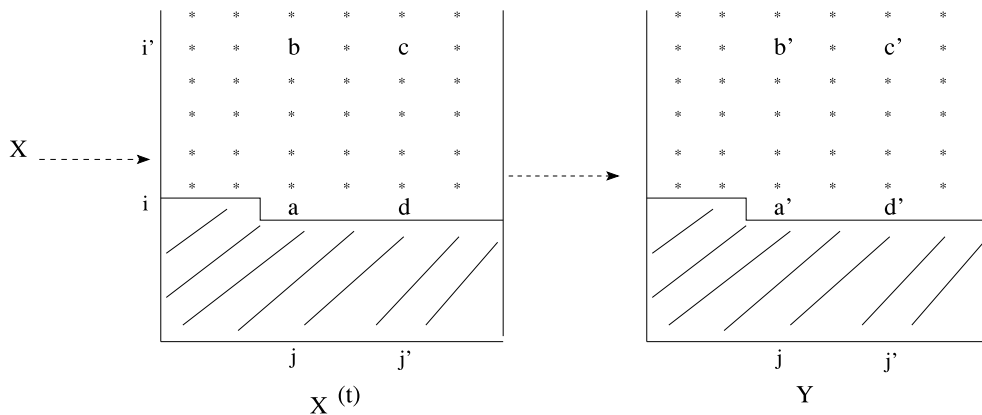
Fig. 2. The hatched zone is the matched zone at time t, where we assume that $i' > i$ and $j' > j$. The picture also assumes that there may be some other contingency tables between X and $X^{(t)}$, and between $X^{(t)}$ and Y.

Theorem 3. *If Ω is the set of $m \times n$ contingency tables, the Diaconis Markov chain on Ω mixes fast and the mixing time $\tau_x(\epsilon)$ is given by*

$$\tau_x(\epsilon) \leq 2c_{max}(mn)^4 \ln(c_{max}\epsilon^{-1}),$$

where $c_{max} - 1$ is the maximal in a cell.

We prove this result by showing that, between any two vertices of $\mathcal{G}$, there is a path, the *canonical path*, that is not 'too long' and where no edge is overloaded. To put this in a more formal setting, we need the following definitions and lemmas.

Let X and Y be two different $m \times n$ contingency tables. If $X_{ij} = Y_{ij}$, we say that the point (i, j) is *matched*, and *unmatched* otherwise. Obviously therefore, to *match* a point (i, j) means to make X_{ij} equal to Y_{ij}. A *path* from X to Y is the sequence of contingency tables $(X = X^{(0)}, X^{(1)}, X^{(2)}, \ldots, X^{(t)}, \ldots, Y)$, where one contingency table is obtained from the preceding one by a single flip on some site. The *canonical path* from X to Y is a path that matches the cells in increasing lexicographic ordering. Since there are many such paths from X to Y, we take the path that uses the flips with the least indices, lexicographically. This makes the canonical path unique. Suppose that at time t, the canonical path has reached the contingency table $X^{(t)}$, corresponding to matching the cell (i, j), and $X_{ij}^{(t)} \neq Y_{ij}$. The *matched zone* is the set of cells (k, l) such that $(k, l) < (i, j)$ (Obviously $X_{kl}^{(t)} = Y_{kl}$ for $(k, l) < (i, j)$). The cell (i, j) is then called the *smallest unmatched cell*. A key observation, given in the lemma that follows, is that once a cell is matched, the canonical path does not unmatch it anymore. This entails that it is possible to match all the mn cells in the lexicographic order, without backtracking.

Lemma 4. *Let (i, j) be the smallest unmatched cell at time t along the canonical path from X to Y. There is a sequence of flips where all the other cells involved in matching (i, j) are greater than (i, j).*

Proof. Let (i, j) be the least unmatched cell at time t, and suppose that $X_{ij}^{(t)} = a$ and $Y_{ij} = a'$, with $|a' - a| = r$, as illustrated in Fig. 2.

If there is no row i' with $i' > i$ (if i is the last row), then $X^{(t)}$ and Y have different column sums on the column j, which is a contradiction. A similar argument holds if there is no column j' such that $j' > j$. Therefore there is a row i' and column j' such that $i' > i$ and $j' > j$. Thus making r successive flips on $[i, i', j, j']$ matches the cell (i, j). We only have to care about the case where some entries on the site $[i, i', j, j']$ become negative before performing all the r flips.

1. $a < a'$. In this case, either (i', j) or (i, j'), or both, may become negative. Suppose that the entry on the point (i', j) becomes negative at the sth flip, with $s < r$. Then there must be other rows k_i, with $k_i > i$ which have together at least $r - s + 1$ positive entries, such that successive flips on the site $[i, k, j, j']$ can complete the

matching. If there is no such row, then all the entries $X_{kj}^{(t+s-1)} = 0$ for all $k > i$. But since all the points (l, j) with $l < i$ are matched, we have c_j in $X^{(t+s-1)}$ is less than c_j in Y, a contradiction. A similar argument holds if the entry in the cell (i, j') became negative.

2. $a > a'$. In this case, (i', j') may become negative. Suppose that the entry on the point (i', j') becomes negative at the sth flip, with $s < r$. Then, there are rows k and column l, with $k > i$ and $l > j$, which have together at least $r - s + 1$ entries such that the successive flips on $[i, k, j, l]$ complete the matching of the cell (i, j). Suppose there is no such row k, or column l, or both, since all the relevant entries are zeros. Thus, either the row sum r_i in $X^{(t+s-1)}$ is less than r_i in Y, or the column sum c_j in $X^{(t+s-1)}$ is less than c_j in Y or both. Again a contradiction. □

Corollary 5. *For every pair of contingency tables X and Y, there is a sequence of flips changing X to Y such that the points in successive matched zones are not affected.*

Corollary 6. *For all contingency tables X and Y, the length of the canonical path from X to Y is at most $(c_{max} - 1)mn$, where $c_{max} - 1$ is the maximal value in a cell.*

Proof. To match the point (i, j) requires at most $c_{max} - 1$ successive flips. Since there are at most mn points to match and there is no backtracking by Corollary 5, the result follows. □

A second key observation, given in Lemma 9, is that the number of canonical paths which pass through an edge e is not too great. We first need the following definition and a lemma of far reaching importance.

Let Ω denote the set of $m \times n$ contingency tables, where r_i and c_j are the sums of row i and column j, respectively, for $1 \le i \le m$, $1 \le j \le n$, and let $X \in \Omega$. A *margin* is either a column or a row sum. Let X_j denote the jth column of X. Suppose that the column X_j contains m entries and the sum of these m entries is c_j, then the number of non-negative integer solutions of the equation $x_1 + x_2 + \cdots + x_m = c_j$ is equal to $\binom{c_j + m - 1}{m - 1}$. We denote by X_{j_l} the lth non-negative integer solution of the equation $x_1 + x_2 + \cdots + x_m = c_j$. (Not to be confused with $X_{j,l}$, the (j, l)th entry of the contingency table X.) Let $Y^{(l)}$ be a contingency table whose jth column is the solution X_{j_l}. (By Theorem 2, starting from X, it is possible to reach any $Y^{(l)}$. We assume that $X = Y^{(1)}$). We denote by $Y^{(l)} \setminus X_{j_l}$ the contingency table obtained by deleting column X_{j_l} from $Y^{(l)}$. Obviously, $Y^{(l)} \setminus X_{j_l}$ is an $m \times (n - 1)$ contingency table whose row sums are $r_i - a_i$, where a_i is the ith entry of column j of $Y^{(l)}$. We now state and prove the following instrumental lemma.

Lemma 7. *Let Ω denote the set of $m \times n$ contingency tables where r_i and c_j are the sums of row i and column j, respectively, for $1 \le i \le m$, $1 \le j \le n$. Let N be the cardinality of Ω. Choose any contingency table $X \in \Omega$ and let X_j be the column or row of X with the smallest margin amongst all the columns and rows of X. Then,*

$$N = N_1' + N_2' + \cdots + N_\sigma' \tag{1}$$

where N_l' is the number of $m \times (n - 1)$ contingency tables that can be reached from $Y^{(l)} \setminus X_{j_l}$ through a sequence of flips, $\sigma = \binom{c_j + m - 1}{m - 1}$ is the number of non-negative integer solutions of the equation $x_1 + x_2 + \cdots + x_m = c_j$, and c_j is the margin of X_j.

Proof. (We insist that X_j may be a row or a column, as long as it has the least margin. But even when it is a row, we still consider it as a column, by just transposing the table.) Consider any contingency table $Y^{(l)} \in \Omega$ and consider its column j. The set Ω is the disjoint union of A and B, where A is the set of contingency tables obtained from $Y^{(l)}$ through a sequence of flips on sites not containing entries of column j, and B is the set of contingency tables obtained from $Y^{(l)}$ through a sequence of flips where at least one site containing entries of column j is involved. The set A is isomorphic to the set of $m \times (n - 1)$ contingency tables that can be reached from $Y^{(l)} \setminus X_{j_l}$ after a sequence of flips. Thus, $|A| = N_l'$.

Elements of the set B can be obtained from $Y^{(l)}$ by performing a single flip in a site containing entries of column j of $Y^{(l)}$ to obtain a contingency table $Y^{(l+1)}$, and, from $Y^{(l+1)}$ perform all the possible flips in sites not

containing entries in column j of $Y^{(l+1)}$. As above, this is isomorphic to the set of $m \times (n-1)$ contingency tables reachable from $Y^{(l+1)} \setminus X_{j,l+1}$. There are N'_{l+1} such contingency tables. Recursively, one may thus obtain all the $N'_1, N'_2, \ldots, N'_\sigma$, starting from $l = 1$, where $\sigma = \binom{c_j+m-1}{m-1}$, is the number of non-negative integer solutions of the equation $x_1 + x_2 + \cdots + x_m = c_j$.

It only remains to show that each of these solutions would yield a valid contingency table in Ω. For a contradiction, let Z_j be a non-negative integer solution of the equation $x_1 + x_2 + \cdots + x_m = c_j$ and there is no contingency table Z such that Z_j is its jth column. Then there is no contingency table in Ω whose jth column can be transformed into Z_j through a sequence of flips. Let a_r and a_s be two entries in Z_j. We may assume that a_s is in the sth position (row) and a_r is in the rth position. There is no contingency table Z in Ω such that Z_j is its jth column only if there is no contingency table $Z^{(1)}$ in Ω whose (r, j)th entry is $a_r + 1$ and the (s, j)th entry is $a_s - 1$, and such that a flip in the site $[r, s, j, j']$ of $Z^{(1)}$ produces Z. This is possible only if, in $Z^{(1)}$, all the entries in row s (apart from a_s) are zeros (so that none can be decreased by 1). Thus either (1), $Z^{(1)}_{s,j} = r_s$, where r_s is the sum of row s, or (2), there is another entry that is not zero in row s.

(1) $Z^{(1)}_{s,j} = r_s$. Either $r_s > c_j$ or not. If $r_s > c_j$, there will be a negative entry in $Z^{(1)}$, which then is not a contingency table. To avoid this, we chose X_j to be the column or row of the least sum amongst all the column and row sums. If not, then

(2) Suppose there is another entry that is not zero in row s. Let this non-zero entry be in column j'. Then a flip on site $[r, s, j, j']$ may not be possible only if the (r, j') entry of $Z^{(1)}$ is equal to the sum of row r, so that it cannot be increased by 1. Thus $Z^{(1)}_{r,j} = 0$.

Therefore, for all contingency tables $Z^{(1)}$ in Ω whose (r, j)th entry is $a_r + 1$ and (s, j)th entry is $a_s - 1$, and all other entries are equal to the corresponding entries of Z_j, we have either $Z^{(1)}_{r,j} = 0$ or $Z^{(1)}_{s,j} = r_s$. Thus $a_s = r_s + 1$ or $a_r = 0 - 1$. In either case, Z_j is not a non-negative integer solution of the equation $x_1 + x_2 + \cdots + x_m = c_j$. This is a contradiction. □

Corollary 8. *Let N be the number of all $m \times n$ contingency tables of fixed row and column sums. The number of contingency tables having k fixed cells (in lexicographic ordering) is at most $N^{\frac{mn-k}{mn}}$.*

Proof. We prove the result by induction on k, the number of cells that are fixed in the lexicographical ordering. Indeed, if $k = 0$, the number of such contingency tables is N. In case $k = mn$, there is only a single contingency table. Hence the formula holds for these two extreme cases.

For the induction, assume that the result holds for k. Let $k + 1$ cells be fixed and let the $(k + 1)th$ cell be on the column j. Let Ω denote the set of contingency tables having $k + 1$ vertices fixed (in lexicographic ordering). We have to show that $|\Omega| \leq N^{\frac{mn-k-1}{mn}}$.

By Lemma 7, and by induction, we have that

$$|\Omega| \leq N'^{\frac{m(n-1)-k}{m(n-1)}}_1 + N'^{\frac{m(n-1)-k}{m(n-1)}}_2 + \cdots + N'^{\frac{m(n-1)-k}{m(n-1)}}_\sigma$$
$$\leq (N'_1 + N'_2 + \cdots + N'_\sigma)^{\frac{m(n-1)-k}{m(n-1)}}$$
$$\leq N^{\frac{m(n-1)-k}{m(n-1)}} \leq N^{\frac{mn-k-1}{mn}}. \quad \square$$

Lemma 9. *If $e = (z, y)$ is an arc of $\mathcal{G}$, then the number of different canonical paths passing through e is at most N, where N is the number of vertices of $\mathcal{G}$.*

Proof. Let $\sigma = (v, \ldots, z, y, \ldots, w)$ be a canonical path passing through the edge $e = (z, y)$, as illustrated in Fig. 3. In the graph $\mathcal{G}$, the edge e represents a flip on a fixed site of the contingency table z, such that making the flip e moves the chain from z to y. Suppose that the z is reached at the kth flip along the canonical path σ, and let $\bar{Z}$ be the matched zone at z. That is, $\bar{Z}$ is the set containing the first k cells, in the lexicographic ordering, i.e., $\bar{Z} = (1, 2, \ldots, k)$. Now, by Lemma 4, the matched zone is preserved along the canonical path. Thus w is the end point of a canonical path passing through e only if w contains $\bar{Z}$. Hence w are contingency tables where the sequence of cells $(1, 2, \ldots, k)$ is
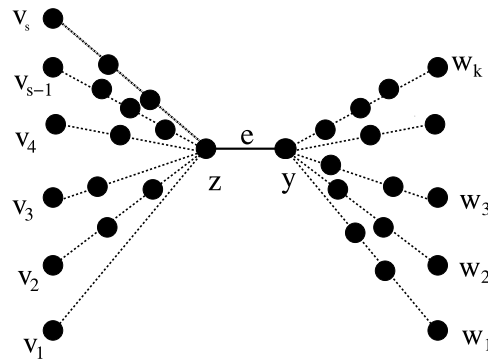
Fig. 3. The set of canonical paths passing through the edge e. Notice that e represents the flip that changes the contingency table Z to the contingency table Y.

fixed, and the sequence of cells $(k + 1, k + 2, \dots, mn)$ may take any entry, subject to being a contingency table of the correct size. The number of such vertices (contingency tables) is thus at most $N^{\frac{mn-k}{mn}}$, by Lemma 7.

Moreover, let $\bar{Z}^c$ denote the complement of $\bar{Z}$. That is $\bar{Z}^c = (k + 1, k + 2, \dots, mn)$. The contingency table v is the start point of the canonical path σ only if v is such that matching the k first cells simultaneously turns the sequence of entries in $(k + 1, k + 2, \dots, mn)$ equal to the corresponding entries in z. Hence, since the canonical path is unique, the sequence of entries in $(1, 2, \dots, k)$ uniquely determines the sequence of entries in $(k + 1, k + 2, \dots, mn)$. Therefore v are contingency tables where the sequence of cells $(k + 1, k + 2, \dots, mn)$ is uniquely determined by the sequence of cells $(1, 2, \dots, k)$. The latter may take any entry, subject to being a contingency table of the right size. The number of such vertices (contingency tables) is thus at most $N^{\frac{k}{mn}}$.

Hence, the number of canonical paths using e is at most $N^{\frac{k}{mn}} N^{\frac{mn-k}{mn}} = N$. □

Now, we are able to prove Theorem 3.

Proof of Theorem 3. In order to prove Theorem 3 by using Theorem 1, we show that there is a flow ϕ such that $\rho(\phi)$ is a polynomial in $m \times n$, the size of contingency tables in Ω. Indeed, if x and y are two vertices of $\mathcal{G}$, let $\hat{p}_{xy}$ denote the canonical path from x to y and let ϕ be a flow consisting of injecting $\pi(x)\pi(y)$ units of flow along $\hat{p}_{xy}$. Then, for all arcs e, we have

$$\phi'(e) = \sum \pi(x)\pi(y),$$

where the sum is over all the pairs $\{x, y\}$ such that $e \in \hat{p}_{xy}$. Since by Lemma 9, there are at most N canonical paths through e, and $\pi(x) = \pi(y) = \frac{1}{N}$, as the distribution π is uniform, we have

$$\phi'(e) \leq N\pi(x)\pi(y) = \frac{1}{N}.$$

Moreover, since $P_{x,y} = \frac{1}{2(mn)^2}$ if there is an edge from x to y, we have $c := c(e) = \pi(x)P_{x,y} = \frac{1}{N(2m^2n^2)}$ is constant for all e. Thus $\rho(\phi) = \frac{max_e\phi'(e)}{c}$.

Now, using Theorem 1, Corollary 6, we have

$$\tau_x(\epsilon) \leq (2m^2n^2)(c_{max}mn)(\ln \frac{1}{\pi(x)} + \ln \frac{1}{\epsilon}).$$

Since an entry may take any value from 0 to $c_{max} - 1$, we have $N \leq (c_{max})^{mn}$. Thus,

$$\tau_x(\epsilon) \leq 2c_{max}(mn)^3(mn \ln c_{max} + \ln \frac{1}{\epsilon}) \leq 2c_{max}(mn)^4 \ln(c_{max}\epsilon^{-1}). □$$

Acknowledgment

References

[1] P. Diaconis, A. Gangolli, Rectangular arrays with fixed margins, in: Proceedings of the Workshop on Markov Chains, Institute of Mathematics and its Applications, University of Minnesota, 1994.

[2] P. Diaconis, B. Efron, Testing for independence in a two-way table, Ann. Statist. 13 (1985) 845–913.

[3] N.J. Goteli, A.M. Ellison, A Primer of Ecological Statistics, Second ed., Sinauer Associates, Inc, 2013.

[4] S. Pirzada, An Introduction To Graph Theory, Universities Press, OrientBlackswan, Hyderabad, India, 2012.

[5] R.A. Brualdi, Matrices of zeroes and ones with fixed row and column sum vectors, Linear Algebra Appl. 33 (1980) 159–231.

[6] G.W. Cobb, Y. Chen, An application of Markov Chains Monte Carlo to community ecology, Amer. Math. Monthly 110 (2003) 265–288.

[7] C. Cooper, M. Dyer, C. Greenhill, Sampling regular graphs and Peer-to-Peer network, Combin. Probab. Comput. 16 (2007) 557–594.

[8] P. Erdös, T.G. Gallai, Graphs with prescribed degrees of vertices (Hungarian), Mat. Lapok 11 (1960) 264–274.

[9] M. Gail, N. Mantel, Counting the number of $r \times c$ contingency tables with fixed margins, J. Amer. Statist. Assoc. 72 (1977) 859–862.

[10] M. Jerrum, A. Sinclair, Approximating the permanent, SIAM J. Comput. 18 (6) (1989) 1149–1178.

[11] R. Kannan, P. Tetali, S. Vempala, Simple Markov-chain algorithms for generating bipartite graphs and tournaments, Random Struct. Algorithms 14 (4) (1999) 293–308.

[12] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow, Lect. Notes Comput. Sci. 583 (1992) 474–487.

[13] R. Kannan, Markov Chains and polynomial time Algorithm, in: Plenary Talk At Proc. of 35th Annual IEEE Symp. on the Foundations of Computer Science, 1994, pp. 656–671.

[14] P.L. Erdös, I. Miklós, Z. Toroczkai, A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs, Electron. J. Combin. 17 (1) (2010) P66.

[15] Y. Chen, P. Diaconis, S. Holmes, J.S. Liu, Sequential Monte Carlo methods for statistical analysis of tables, J. Amer. Statist. Assoc. 100 (2005) 109–120.

[16] F.R.K. Chung, R.L. Graham, S.T. Yau, On sampling with Markov chains, Random Struct. Algorithms 9 (1996) 55–77.

[17] D.A. Levin, Y. Peres, E.L. Wilmer, Markov Chains and Mixing Times, Amer. Math. Soc, 2006.