Random Generation of $2 \times n$ Contingency Tables

Diane Hernek
Department of Mathematics, UCLA, Los Angeles, CA 90095;
e-mail: dhernek@math.ucla.edu

Received 11 March 1997; accepted 19 May 1998

ABSTRACT: Let $r=(r_1, r_2)$ and $c=(c_1, \dots, c_n)$ be positive integer partitions of N. Let Σ_{rc} denote the set of all $2 \times n$ arrays of nonnegative integers whose ith row sums to r_i and jth column sums to c_j. We consider the problem of randomly generating an element from the uniform distribution over Σ_{rc}. This problem arises in statistics where random samples are used to decide whether two attributes are independent. In this paper, we present a Markov chain Monte Carlo algorithm for this problem and give the first general polynomial bounds on its running time. © 1998 John Wiley & Sons, Inc. Random Struct. Alg., 13, 71–79, 1998

1. INTRODUCTION

Let $r=(r_1, \dots, r_m)$ and $c=(c_1, \dots, c_n)$ be positive integer partitions of N. Let Σ_{rc} be the set of all $m \times n$ arrays of nonnegative integers whose ith row sums to r_i and jth column sums to c_j. That is, Σ_{rc} is the set of all $m \times n$ arrays $\{T_{ij}\}$ of nonnegative integers such that $\sum_{j=1}^{n} T_{ij} = r_i$, for $1 \le i \le m$, and $\sum_{i=1}^{m} T_{ij} = c_j$, for $1 \le j \le n$. The set Σ_{rc} is always nonempty and an element of Σ_{rc} can be constructed efficiently using a greedy method. This paper is concerned with the more difficult problem of randomly generating an element from the uniform distribution over Σ_{rc}, given r and c as input. This problem can be solved quite simply by systematically enumerating all elements of Σ_{rc}. In general, however, the set Σ_{rc} will be very large and we would like to solve this problem efficiently. More precisely, while the number of arrays may be exponential n m, n, and N, we would like to sample in time polynomial in m, n, and N.

This problem arises in statistics where such arrays are called contingency tables. Consider a scenario where N subjects are categorized into a contingency table T according to two attributes. For example, in Figure 1 [2], a random sample of $N = 500$ persons is categorized by political affiliation and attitude toward an energy-rationing program. Data is often analyzed under the assumption that the two attributes are independent; that is, that the joint distribution is uniquely determined by the marginal probabilities. More precisely, if $p_{ij} = T_{ij}/N$, the independence assumption asserts that $p_{ij} = p_{i.}p_{.j}$, where $p_{i.} = \Sigma_j p_{ij}$ and $p_{.j} = \Sigma_i p_{ij}$. One test that is often used as a measure of independence is the χ^2 ("chi-square") statistic, defined below,

$$\chi^2 = \sum_{i,j} \frac{\left(T_{ij} - \frac{r_i c_j}{N}\right)^2}{\frac{r_i c_j}{N}}.$$

Under the independence assumption, conditioning on the row and column sums, c_j/N is the fraction of the r_i subjects in row i that we expect to see in cell i, j of the table. Intuitively, a small χ^2 value supports the independence hypothesis.

An alternative approach [3] is to assume that the table was generated from the uniform distribution over Σ_{rc}. A typical test of this hypothesis asks what fraction of tables in Σ_{rc} have χ^2 smaller than t, as t varies. Exact calibration can be done by systematically enumerating all tables in Σ_{rc}. When Σ_{rc} is very large, enumeration may be impractical or even impossible. In this situation one would like to estimate the moments of the χ^2 statistic, over the set Σ_{rc}, by randomly sampling from Σ_{rc}.

The following random walk method for sampling an element of Σ_{rc} has been proposed (see, for example, [4]) and is used extensively in statistical applications. Start with an arbitrary element T of Σ_{rc} and construct T' according to the following randomized step. Select a pair of rows $1 \le i_1 < j_1 \le m$ and a pair of columns $1 \le i_2 < j_2 \le n$ at random and flip a fair coin. If the outcome of the coin is heads, T' is defined by adding one to $T_{i_1 i_2}$ and $T_{j_1 j_2}$, subtracting one from $T_{i_1 j_2}$ and $T_{j_1 i_2}$, and leaving all other cells of T unchanged. If the outcome of the coin is tails, T' is defined by subtracting one from $T_{i_1 i_2}$ and $T_{j_1 j_2}$, adding one to $T_{i_1 j_2}$ and $T_{j_1 i_2}$, and leaving all other cells of T unchanged. If all cells of T' are nonnegative, the walk moves to T', otherwise the walk remains at T. It has been proposed that, after a moderate number of repetitions of this step, the distribution of the current state is approximately uniform over Σ_{rc} and, therefore, can be used as a sample from the uniform distribution. In this paper, we use a coupling argument to prove that in case of two rows this Markov chain mixes in polynomial time. In particular, we show that the number of walk steps required to obtain a random $2 \times n$ array is at most quadratic in n and N. It is an open question to determine whether the chain mixes in polynomial time for arrays with more than two rows.

	Favor	Indifferent	Opposed	TOTAL
Democrat	138	83	64	285
Republican	64	67	84	215
TOTAL	202	150	148	500

Fig. 1. Contingency table for political affiliation and opinion.

It is worth pointing out that in the $2 \times n$ case the counting problem can be solved in polynomial time using a dynamic programming approach. That is, it is possible to compute $|\Sigma_{rc}|$ exactly, in time polynomial in n and N, given $r = (r_1, r_2)$ and $c = (c_1, \dots, c_n)$ as input. Using this counting procedure as a subroutine, it is then possible to generate an array from the uniform distribution over Σ_{rc} in time polynomial in n and N. In practice, however, the Markov chain considered in this paper is used to generate samples from Σ_{rc}. The contribution of this paper is to analyze this specific approach. Moreover, while the dynamic programming approach will not generalize to sampling arrays with more than two rows, the hope is that the Markov chain approach will.

In related work, Diaconis and Saloff-Coste [5, 6] have shown that the present chain mixes in polynomial time in N when both m and n are fixed. In contrast, Dyer, Kannan, and Mount [8] use a different Markov chain for sampling $m \times n$ arrays in the case of sufficiently large row and column sums. Their algorithm produces a sample in polynomial time provided that $r_i = \Omega(mn^2)$ and $c_j = \Omega(m^2 n)$.

2. THE MARKOV CHAIN

In this section we give a more formal definition of the Markov chain corresponding to the random walk described above. The state space Ω of the Markov chain is Σ_{rc}, the set of all $2 \times n$ arrays of nonnegative integers with row sums r_1, r_2 and column sums $c_1, \dots, c_n$. The transitions of the Markov chain are defined as follows. Let T be the array corresponding to the current state. The process moves to a new state T' according to the following rule:

1. With probability $\frac{1}{2}$, self-loop; that is, set $T' = T$.

2. With the remaining $\frac{1}{2}$ probability, choose a random pair of indices $1 \le i < j \le n$.

(a) With probability $\frac{1}{2}$, let T' be defined as follows,

$$\begin{aligned} T'_{1i} &= T_{1i} + 1, \\ T'_{2i} &= T_{2i} - 1, \\ T'_{1j} &= T_{1j} - 1, \\ T'_{2j} &= T_{2j} + 1, \\ T'_{kl} &= T_{kl}, \quad \text{for } l \ne i \quad \text{and} \quad l \ne j. \end{aligned}$$

If T'_{2i} or T'_{1j} is -1, reject this move and set $T' = T$. We call this a plus move.

(b) With probability $\frac{1}{2}$, let T' be defined as follows,

$$\begin{aligned} T'_{1i} &= T_{1i} - 1, \\ T'_{2i} &= T_{2i} + 1, \\ T'_{1j} &= T_{1j} + 1, \\ T'_{2j} &= T_{2j} - 1, \\ T'_{kl} &= T_{kl}, \quad \text{for } l \ne i \quad \text{and} \quad l \ne j. \end{aligned}$$

If T'_{1i} or T'_{2j} is -1, reject this move and set $T' = T$. We call this a minus move.

This Markov chain is aperiodic and irreducible, hence it is ergodic. Aperiodicity follows from the fact that there is a self-loop with nonzero probability at every state. In addition, the Markov chain is irreducible; that is, for any pair of states X, Y, there is a sequence of moves that takes X to Y. We will demonstrate a shortest path that fixes the columns of X to agree with Y in order from 1 to n. Let $\Phi = \Phi(X,Y) = \sum_{i=1}^{n} |X_{1i} - Y_{1i}|$. Note that Φ is necessarily even. We will exhibit a (shortest) path of length $\Phi/2$ from X to Y, by induction. If $\Phi = 0$, then $X = Y$ and there is a path of length 0. If $\Phi > 0$, let i be the minimum such that $X_{1i} \neq Y_{1i}$. Without loss of generality, assume that $X_{1i} < Y_{1i}$. Let $j > i$ be the minimum such that $X_{1j} > Y_{1j}$. Since the first rows of X and Y both sum to r_1, such a j must exist. A plus move on X using the i and j specified reduces Φ by two in one step. By induction, there is a path of length $\Phi/2$ connecting X to Y.

Ergodicity implies the existence of a unique stationary distribution π. For $X \neq Y$, either $P(X,Y) = P(Y,X) = 0$ or $P(X,Y) = P(Y,X) = 1/[n(n-1)]$. This is because plus and minus moves with the same pair of columns reverse each other. Hence, this Markov chain is reversible, satisfying $\pi(X)P(X,Y) = \pi(Y)P(Y,X)$, with respect to the uniform distribution π, so the stationary distribution is uniform.

3. BOUNDING THE MIXING TIME

The previous section establishes that the Markov chain, starting from any initial state, eventually converges to the uniform distribution on its state space. The main contribution of this paper is to show that the number of steps needed for the Markov chain to approach its stationary distribution is small relative to the size of its state space. More precisely, we shall show that the Markov chain approaches stationarity in time polynomial in n, the number of columns, and $R = \min\{r_1, r_2\}$, the smallest row sum, even though, in general, the number of states of the Markov chain may be exponential in these quantities. The proof uses a coupling argument.

Let the random variable X_t denote the current state of the Markov chain at time t. Let X be any state of the Markov chain and suppose that $X_0 = X$. For any state Y, let $P_X^t(Y)$ be the probability that $X_t = Y$ given that $X_0 = X$; that is, let P_X^t be the probability distribution of the current state at time t, given that the random walk began at state X. The distance of the probability distribution at time t from the stationary distribution is measured using total variation distance, $\|P_X^t - \pi\|$, defined below,

$$\|P_X^t - \pi\| = \max_{A \subset \Omega} \left\{P_X^t(A) - \pi(A)\right\} = \frac{1}{2} \sum_{Y \in \Omega} |P_X^t(Y) - \pi(Y)| < 1.$$

For any $0 < \epsilon < 1$, the mixing time $\tau(\epsilon)$ is given by $\tau(\epsilon) = \max_X \inf\{t: \|P_X^t - \pi\| \leq \epsilon\}$. In words, $\tau(\epsilon)$ is the number of steps of a random walk needed for the probability distribution of the current state to be within ϵ of the stationary distribution π, maximized over all initial states.

The mixing time is bounded using the method of coupling. In this method, a stochastic process $\{(X_t, Y_t)\}$ is defined on the state space $\Omega \times \Omega$. In this process, X_t and Y_t are coordinated, however, viewed in isolation, the processes $\{X_t\}$ and $\{Y_t\}$ each evolve according to the original Markov chain. Moreover, once the two processes get into the same state, they remain in the same state forever. That is, for all t, if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$. The idea is to coordinate the two processes in such a way that X_t and Y_t tend to grow closer together according to some suitably defined measure of distance. Let $T(X,Y) = \inf\{t\colon X_t = Y_t \mid X_0 = X,\ Y_0 = Y\}$. The process $\{(X_t, Y_t)\}$ is called a coupling and $T(X,Y)$ is called a coupling time. Let $T = \max_{X,Y \in \Omega} \mathrm{E}[T(X,Y)]$ be the maximum expected coupling time. The relationship between the maximum expected coupling time and the mixing time of Markov chains is well understood (see, for example, Aldous [1]). The following theorem gives an upper bound on the mixing time in terms of maximum expected coupling time.

Theorem 1 (Aldous). *For any $0 < \epsilon < 1$, the mixing time $\tau(\epsilon)$ is related to the maximum expected coupling time T by the inequality $\tau(\epsilon) \le 2eT(1 + \ln \epsilon^{-1})$.*

In the next subsection, we shall construct a coupling and prove that with this coupling T is polynomial in n and N. This combined with Theorem 1 will prove that the Markov chain mixes in time polynomial in n, N, and $\ln \epsilon^{-1}$.

3.1. The Coupling

Let X and Y be any pair of states. We will measure the distance between X and Y using the potential function defined below,

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

Note that $0 \le \Phi \le 2R$, where $R = \min\{r_1, r_2\}$, and that $\Phi = 0$ if and only if $X = Y$. In the coupled process, a single pair of indices $1 \le i < j \le n$ is selected at random. There are four cases (up to symmetry) to consider when defining the transition probabilities of state (X,Y) in the coupling.

1. $X_{1i} < Y_{1i}$ and $X_{1j} > Y_{1j}$
  (a) With probability $\frac{1}{4}$, perform the self-loop on both X and Y.
  (b) With probability $\frac{1}{4}$, perform the self-loop on X and the minus move on Y.
  (c) With probability $\frac{1}{4}$, perform the plus move on X and the self-loop on Y.
  (d) With probability $\frac{1}{4}$, perform the minus move on X and the plus move on Y.
  (The case where $X_{1i} > Y_{1i}$ and $X_{1j} < Y_{1j}$ is symmetric with the roles of the plus move and the minus move interchanged.)

2. $X_{1i} < Y_{1i}$ and $X_{1j} < Y_{1j}$
  (a) With probability $\frac{1}{4}$, perform the self-loop on X and the plus move on Y.
  (b) With probability $\frac{1}{4}$, perform the self-loop on X and the minus move on Y.
  (c) With probability $\frac{1}{4}$, perform the plus move on X and the self-loop on Y.
  (d) With probability $\frac{1}{4}$, perform the minus move on X and the self-loop on Y.
  (This move is symmetric with respect to X and Y so the case where $X_{1i} > Y_{1i}$ and $X_{1j} > Y_{1j}$ is identical.)

3. $X_{1i} < Y_{1i}$ and $X_{1j} = Y_{1j}$
   (a) With probability $\frac{1}{2}$, perform the self-loop on both X and Y.
   (b) With probability $\frac{1}{4}$, perform the plus move on both X and Y.
   (c) With probability $\frac{1}{4}$, perform the minus move on both X and Y.

   (This move is symmetric with respect to X and Y and i and j so there are three other cases whose treatment is identical.)

4. $X_{1i} = Y_{1i}$ and $X_{1j} = Y_{1j}$
   (a) With probability $\frac{1}{2}$, perform the self-loop on both X and Y.
   (b) With probability $\frac{1}{4}$, perform the plus move on both X and Y.
   (c) With probability $\frac{1}{4}$, perform the minus move on both X and Y.

This joint distribution is a valid coupling since $\{X_t\}$ and $\{Y_t\}$ are each faithful to the original Markov chain and, by the definition of case 4, if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.

3.2. The Analysis

Let $\Phi(t) = \Phi(X_t, Y_t)$ be the potential at time t and, for $t \geq 1$, let $\Delta(t) = \Phi(t) - \Phi(t-1)$ be the change in potential at time t. Our aim is to bound $T = \max_{X,Y \in \Omega} \mathrm{E}[T(X,Y)]$, where $T(X,Y) = \inf\{t\colon \Phi(t) = 0 \mid X_0 = X, Y_0 = Y\}$. The key is to show that, for any pair of states, in one step of the coupled process the expected change in potential is nonpositive. This statement is formalized in the following lemma.

Lemma 1. *If* $\Phi(t) > 0$, *then* $\mathrm{E}[\Delta(t+1) \mid (X_t, Y_t)] \leq 0$ *and* $\mathrm{E}[\Delta(t+1)^2 \mid (X_t, Y_t)] \geq V$, *where* $V = 4/[n(n-1)]$.

Proof. We consider the change in potential Φ in each of the four cases of the coupling. To simplify notation we use X to denote X_t and Y to denote Y_t in this proof.

1. Since $X_{1i} < Y_{1i}$ and $X_{1j} > Y_{1j}$, we know that $X_{2i} > Y_{2i}$ and $X_{2j} < Y_{2j}$. The self-loop on both X and Y results in no change in potential. The minus move on Y is never rejected since it decrements $Y_{1i} > X_{1i} \geq 0$ and $Y_{2j} > X_{2j} \geq 0$. Hence, this move always reduces the potential Φ by two. Similarly, the plus move on X is never rejected since it decrements $X_{1j} > Y_{1j} \geq 0$ and $X_{2i} > Y_{2i} \geq 0$. Hence, this move always reduces the potential Φ by two. The minus move on X or the plus move on Y may be rejected. In the worst case, both moves are performed resulting in an increase by four in total potential so, in this case, the change is between zero and four. Averaging, we obtain that $\mathrm{E}[\Delta(t+1) \mid X_{1i} < Y_{1i}, X_{1j} > Y_{1j}] \leq \frac{1}{4}(0 - 2 - 2 + 4) = 0$ and $\mathrm{E}[\Delta(t+1)^2 \mid X_{1i} < Y_{1i}, X_{1j} > Y_{1j}] \geq \frac{1}{4}(0^2 + (-2)^2 + (-2)^2 + 0^2) = 2$.

2. Since $X_{1i} < Y_{1i}$ and $X_{1j} < Y_{1j}$, we know that $X_{2i} > Y_{2i}$ and $X_{2j} > Y_{2j}$. In this case, plus and minus moves are always paired with a self-loop in the other process. Any of the plus or minus moves may be rejected, resulting in no change in potential. Moreover, even if a move is performed, there is no change in potential. For example, if we perform the plus move on X, $Y_{1i} - X_{1i}$ decreases by one and $Y_{1j} - X_{1j}$ increases by one.

3. Since $X_{1i} < Y_{1i}$ and $X_{1j} = Y_{1j}$, we know that $X_{2i} > Y_{2i}$ and $X_{2j} = Y_{2j}$. The self-loop does not change the potential. The plus move decrements X_{1j}, Y_{1j}, X_{2i}, Y_{2i}, where $X_{1j} = Y_{1j}$ and $X_{2i} > Y_{2i} \geq 0$. It follows that the plus move may be performed on both X and Y, rejected on both X and Y, or performed on X and rejected on Y. If the move is performed or rejected in both X and Y, the change in potential is zero. If the move is performed on X and not Y, the change in potential is zero since $Y_{1i} - X_{1i}$ decreases by one and $Y_{1j} - X_{1j}$ increases by one. Similarly, the minus move decrements X_{1i}, Y_{1i}, X_{2j}, Y_{2j}, where $X_{2j} = Y_{2j}$ and $0 \leq X_{1i} < Y_{1i}$. It follows that the minus move may be performed on both X and Y, rejected on both X and Y, or performed on Y and rejected on X. If the move is performed or rejected in both X and Y, the change in potential is zero. If the move is performed on Y and not X, the change in potential is zero since $Y_{1i} - X_{1i}$ decreases by one and $Y_{1j} - X_{1j}$ increases by one.
4. Since $X_{1i} = Y_{1i}$ and $X_{1j} = Y_{1j}$, we know that $X_{2i} = Y_{2i}$ and $X_{2j} = Y_{2j}$, so the same move is performed on both X and Y or rejected on both X and Y and the change in potential is zero.

Since $\Phi(t) > 0$, there is at least one pair of columns $i < j$ satisfying case 1 of the coupling and we choose this pair with probability $2/[n(n-1)]$. Hence, $\mathrm{E}[\Delta(t+1) \mid (X_t, Y_t)] \leq 0$ and $\mathrm{E}[\Delta(t+1)^2 \mid (X_t, Y_t)] \geq 4/[n(n-1)]$. ∎

To obtain the bound on T, we also need a bound on the tail of the distribution of $T(X,Y)$. This is given in the following lemma.

Lemma 2. *For any pair of states X, Y, and any $t > 0$, $P(T(X,Y) > t) \leq (1-p)^{\lfloor t/R \rfloor}$, where $0 < p = p(n,N) \leq 1$.*

Proof. Let X', Y' be any pair of states and let $\Phi = \Phi(X', Y')$. Recall that Φ is an even integer in the range $[0, 2R]$. We claim that there is a sequence of $\Phi/2$ moves in the coupled process that reduces Φ to zero. The proof is by induction on Φ. If $\Phi = 0$, the claim is trivially true. If $\Phi > 0$, let i be the minimum such that $X'_{1i} \neq Y'_{1i}$. Without loss of generality, assume that $X'_{1i} < Y'_{1i}$. Let $j > i$ be the minimum such that $X'_{1j} > Y'_{1j}$. Either of moves 1(b) or 1(c) of the coupling with the i and j specified will reduce Φ by two in one step. By induction, there is a sequence of $\Phi/2$ moves that reduces Φ to zero. Moreover, this sequence is chosen with probability at least p, where $p = p(n,N) > 0$. Hence, for all t', $P(\Phi(t'+R) = 0 \mid (X_{t'}, Y_{t'})) \geq p$. Thus, $P(T(X,Y) > t) \leq (1-p)^{\lfloor t/R \rfloor}$. ∎

With Lemmas 1 and 2 in place, we can now complete the bound on the mixing time by bounding the expected time until Φ is zero. The proof is similar to ones used to obtain bounds on the hitting time for simple random walks.

Theorem 2. *The Markov chain converges in time quadratic in n, the number of columns, and R, the smallest row sum. More precisely,*

$$\tau(\epsilon) \leq 2e(1 + \ln \epsilon^{-1})\Phi(0)(2B - \Phi(0))/V,$$

where $B = 2R$ and $V = 4/[n(n-1)]$.

Proof. Define a stochastic process $\{Z(t)\}$ as follows,

$$Z(t) = \begin{cases} \Phi(t)^2 - 2B\Phi(t) - Vt, & \text{if } t = 0 \text{ or } \Phi(t-1) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $B = 2R$ and $V = 4\ [n(n-1)]$. Note that $Z(t) \leq 0$, for all t. We can verify that, for all $t \geq 0$, $Z(t)$ satisfies (1) $\mathrm{E}[\,|Z(t)|\,] < \infty$, and (2) $\mathrm{E}[Z(t+1) \mid (X_0, Y_0), \dots, (X_t, Y_t)] \geq Z(t)$. Thus, the sequence $\{Z(t)\}$ is a submartingale with respect to the process $\{(X_t, Y_t)\}$. For condition (1), observe that $|Z(t)|$ is, by definition, finite. For condition (2), if $\Phi(t) = 0$, then $Z(t+1) = 0$ and the condition is satisfied. If $\Phi(t) > 0$, then

$$\begin{aligned} Z(t+1) &= \Phi(t+1)^2 - 2B\Phi(t+1) - V(t+1) \\ &= (\Phi(t) + \Delta(t+1))^2 - 2B(\Phi(t) + \Delta(t+1)) - V(t+1) \\ &= Z(t) + 2\Delta(t+1)(\Phi(t) - B) + \Delta(t+1)^2 - V. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathrm{E}[Z(t+1) \mid (X_0, Y_0), \dots, (X_t, Y_t)] \\ &\quad = \mathrm{E}[Z(t+1) \mid (X_t, Y_t)] \\ &\quad = Z(t) + 2\mathrm{E}[\Delta(t+1) \mid (X_t, Y_t)](\Phi(t) - B) + \left(\mathrm{E}\left[\Delta(t+1)^2 \mid (X_t, Y_t)\right] - V\right) \\ &\quad \geq Z(t), \text{ by Lemma 1.} \end{aligned}$$

Let X and Y be any pair of states. The coupling time $T = T(X, Y) = \inf\{t: \Phi(t) = 0\}$ is a stopping time for $\{Z(t)\}$. Using Lemma 2, it is straightforward to check that (1) $P(T < \infty) = 1$ (2) $\mathrm{E}[\,|Z(T)|\,] < \infty$, and (3) $\mathrm{E}[Z_t I_{\{T > t\}}] \to 0$ as $t \to \infty$. Hence, by the optional stopping theorem for submartingales (see, for example, [7]), $\mathrm{E}[Z(T)] \geq \mathrm{E}[Z(0)]$. Since $\mathrm{E}[Z(T)] = -V\mathrm{E}[T]$ and $\mathrm{E}[Z(0)] = \Phi(0)^2 - 2B\Phi(0)$, we get that the expected coupling time $\mathrm{E}[T] \leq \Phi(0)(2B - \Phi(0))/V$, which is at most quadratic in R and n. The bound on $\tau(\epsilon)$ now follows directly from Theorem 1. ■

4. CONCLUDING REMARKS

The Markov chain described in this paper is used in statistics to uniformly generate $m \times n$ arrays with row sums $r = (r_1, \dots, r_m)$ and column sums $c = (c_1, \dots, c_n)$. Empirical evidence suggests that the Markov chain mixes in time polynomial in m, n, and N, however, a rigorous proof has remained elusive. Indeed, a rigorous result of this type would constitute a major breakthrough. In this paper, we have used a coupling argument to prove an explicit polynomial bound on the mixing time for the case $m = 2$. It remains to be seen whether the proof can be generalized to larger values of m.

REFERENCES

[1] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, in *Séminaire de Probabilités XVII*, Springer Lecture Notes in Mathematics, A. Dold and B. Eckmann, Eds., 1982, pp. 243–297.

[2] G. K. Bhattacharyya and R. A. Johnson, *Statistical Concepts and Methods*, John Wiley & Sons, New York, 1977.

[3] P. Diaconis and B. Effron, Testing for independent in a two-way table: new interpretations of the chi-square statistic (with discussion), *Ann. Statist.* **13**, 845–913 (1985).

[4] P. Diaconis and A. Gangolli, Rectangular arrays with fixed margins, in *Discrete Probability and Algorithms*, D. Aldous, P. Diaconis, J. Spencer, and J. M. Steele, Eds., Springer-Verlag, Berlin/New York, 1995, pp. 15–41.

[5] P. Diaconis and L. Saloff-Coste, Nash inequalities for Markov chains, *J. of Theoret. Probab.* **9**(2), 459–510 (1996).

[6] P. Diaconis and L. Saloff-Coste, Random walk on contingency tables with fixed row and column sums, unpublished manuscript.

[7] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, Belmont, California, 1996.

[8] M. Dyer, R. Kannan, and J. Mount, Sampling contingency tables, *Random Structures and Algorithms*, **10**(4), 487–506 (1997).