Improved Bounds for Sampling Contingency Tables

Ben J. Morris*
Statistics Department, University of California, Berkeley, CA 94720-3860

Received 6 November 2000; accepted 7 February 2001

DOI 10.1002/rsa.10049

ABSTRACT: We study the problem of sampling contingency tables (nonnegative integer matrices with specified row and column sums) uniformly at random. We give an algorithm which runs in polynomial time provided that the row sums r_i and the column sums c_j satisfy $r_i = \Omega(n^{3/2} m \log m)$, and $c_j = \Omega(m^{3/2} n \log n)$. This algorithm is based on a reduction to continuous sampling from a convex set. The same approach was taken by Dyer, Kannan, and Mount in previous work. However, the algorithm we present is simpler and has weaker requirements on the row and column sums.
© 2002 Wiley Periodicals, Inc. Random Struct. Alg., 21: 135–146, 2002

1. INTRODUCTION

1.1. The Problem

Given positive integer vectors $r = (r_i)_{i=1}^m$ and $c = (c_j)_{j=1}^n$ with $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j$ and $m, n \geq 2$, let $I(r, c)$ denote the set of nonnegative integer $m \times n$ matrices with row sums $r_1, \ldots, r_m$ and column sums $c_1, \ldots, c_n$. In this paper, we consider the problem of generating an element of $I(r, c)$ uniformly at random.

1.2. Motivation

We will now give a brief sketch of how this problem arises in Statistics. The interested reader should consult [1] for a comprehensive account.

*Supported by an NSF Postdoctoral Fellowship.
© 2002 Wiley Periodicals, Inc.

Suppose that we perform an experiment in which N independent samples are taken and classified according to two characteristics A and B, which take the values $1, \ldots, m$, and $1, \ldots, n$, respectively. For example, A might classify each subject's blood type and B might measure cholesterol level. We then assemble the results in an $m \times n$ matrix X such that X_{ij} is equal to the number of samples having $A = i$ and $B = j$. Such a matrix is called a *contingency table.* We will be interested in measuring the amount of dependence between the two variables in a contingency table. A traditional way to quantify this dependence is via the chi-squared statistic:

$$\chi^2(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\left(X_{ij} - \frac{r_i c_j}{N}\right)^2}{\frac{r_i c_j}{N}}.$$

This is the sum of (Observed $-$ Expected)^2/Expected, where "Expected" refers to the quantity $\frac{r_i c_j}{N}$, which is the number of observations we would expect to see in cell ij if A and B were independent with $\Pr(A=i)=\frac{r_i}{N}$ and $\Pr(B=j)=\frac{c_j}{N}$. The *p-value* is defined as the probability that N samples from a distribution with independent row and column variables (i.e., independent A and B) would give a chi-squared statistic which is at least as large as the observed value. Thus, when the p-value is small, there is evidence that the variables are dependent.

Consider the following simple examples with $m = n = 2$. Let

$$X_1 = \begin{bmatrix} 10 & 20 \\ 40 & 30 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} 20 & 40 \\ 80 & 60 \end{bmatrix}.$$

We have $\chi^2(X_1) = 4.76$, which gives a p-value of about .03, and $\chi^2(X_2) = 9.52$, for a p-value of about .002. Thus, comparing p-values we would conclude that there is more dependence in the second data set than in the first. However, both data sets appear to have come from a similar underlying distribution, namely, something like

$$\begin{bmatrix} .1 & .2 \\ .4 & .3 \end{bmatrix}.$$

The second p-value is smaller only because the sample size is larger. This illustrates a problem with the traditional approach of using the p-value alone to measure dependence. While the p-value is useful for determining the *existence* of dependence, it should not be used to measure the *amount* of dependence. Thus it is unwise to compare a p-value or chi-squared statistic from one experiment with that of another. This and other considerations led Diaconis and Efron [1] to propose the following statistic:

$$T(X) = \frac{\left|\{X' \in I(r, c) : \chi^2(X') < \chi^2(X)\}\right|}{\left|I(r, c)\right|},$$

where r and c are the row and column sums, respectively, of X. $T(X)$ is the fraction of all contingency tables in $I(r, c)$ which have a smaller chi-squared statistic than the observed value. $T(X)$ is not highly sensitive to sample size, and is thus a better measure of dependence. In the examples above, we have $T(X_1) = \frac{9}{31} \approx \frac{19}{61} = T(X_2)$. Of course, when the contingency tables are large we cannot always calculate $T(X)$ exactly. This explains why we want an algorithm to sample uniformly from $I(r, c)$; given such an algorithm, we can estimate $T(X)$ in the following way:

1. Take a large number of independent samples from $I(r, c)$.
2. Compute the fraction of samples X' for which $\chi^2(X') < \chi^2(X)$.

1.3. Results

Our method of sampling will rely on the fact that $I(r, c)$, when viewed as a subset of $\mathbf{R}^{mn}$, is the intersection of a continuous convex set and the integer lattice. Specifically, we have $I(r, c) = \mathbf{Z}^{mn} \cap K$, where K is the set of nonnegative *real* matrices with the given row and column sums. A number of (random walk-based) polynomial-time algorithms have been developed in recent years for sampling (almost) uniformly from a convex set (see [2], [10], [7], and [9]). We will solve our discrete sampling problem by reducing it to a continuous one. Given a convex set K' which contains $I(r, c)$, we can generate a random sample from $I(r, c)$ using the following algorithm:

1. Generate a random point Y in K', and "round" it to an integer point Z.
2. If $Z \notin I(r, c)$, repeat.

The two main ingredients in this sampling technique are the convex set K' and the rounding method. The choice of K' is a delicate matter. We require that the distribution of the final sample be almost uniform. Thus, as X varies over $I(r, c)$, we require that vol(X) is nearly constant, where vol(X) denotes the volume of points in K' which round to X. Thus we must make K' sufficiently large. (We could not, for example, naively set K' equal to K, since under any reasonable rounding procedure this would result in a distribution with too little mass on matrices with a large number of zero entries.)

However, there is a tradeoff between the uniformity of the distribution and the running time of the algorithm. As K' becomes larger, it becomes more likely that each sample Z will fall outside of $I(r, c)$. Thus, if we make K' too large, then the expected number of trials taken before $Z \in I(r, c)$ will be too high.

What we want therefore is a convex set K' which is a good continuous approximation to the discrete set $I(r, c)$. Now, since the convex set K is a polytope, it can be defined in terms of bounding hyperplanes. We can therefore define K' in terms of another set of hyperplanes which are parallel to the original ones, only spread farther apart to ensure that vol(X) ≈ 1 for all $X \in I(r, c)$.

The approach we have just described has formed the basis for previous work on sampling contingency tables. Dyer, Kannan, and Mount [4] describe an algorithm similar to the one above and show that it samples (almost) uniformly from $I(r, c)$ in polynomial time, provided that the row and column sums satisfy $r_i = \Omega(n^2 m)$ for all $1 \le i \le m$, and $c_j = \Omega(m^2 n)$ for all $1 \le j \le n$. These were essentially the best known bounds for general m and n (although when $m = 2$, there is a polynomial-time algorithm for any values of r and c; see [3]). In this paper, we will show that the requirements can be loosened to $r_i = \Omega(n^{3/2} m \log m)$ and $c_j = \Omega(m^{3/2} n \log n)$. We accomplish this using the following method for rounding. For $Y \in K'$, we round Y to the integer matrix Z which has $|Z_{ij} - Y_{ij}| \le 1/2$ for all $i < m$, $j < n$, and has the appropriate row and column sums. (Since the samples from K' are continuous there is a unique such Z with probability 1.) This rounding method is quite simple, and it allows us to prove easy bounds on vol(X) for $X \in I(r, c)$. In turn, this allows us to determine the best choice for the convex set K', and leads to the improved requirements on the row and column sums.

The contingency tables problem is a special case of the problem of sampling from the set of integer points in a polytope. This is a class of problems that was studied by Kannan and Vempala in [8], where they give conditions on the polytope which guarantee a polynomial-time algorithm. (When they apply their results to the contingency tables problem, they improve on the row and column sum requirements given in [4], but only by logarithmic factors.) We believe that the techniques in this paper may extend to other problems of this general type.

2. THE ALGORITHM

We will now describe the algorithm for sampling from $I(r, c)$ in detail. Two things are needed to specify the algorithm. First, we need to define the convex set K' from which we perform continuous sampling. Second, we need to describe the rounding method which maps elements of K' to integer lattice points. Since the choice of K' will depend on the rounding method, we will discuss the rounding method first.

We will assume, without loss of generality, that $m \le n$ and that r_m is the largest row sum. Note that any $m \times n$ matrix X whose row and column sums are fixed can be completely specified by $(X_{ij})_{i<m, j<n}$. It will be helpful to think of matrices in $I(r, c)$ as elements of $\mathbf{R}^{(m-1)(n-1)}$ which are indexed by $\{(i, j) : i < m, j < n\}$. Thus, we define $I(r, c)$ as the set of $(m - 1) \times (n - 1)$ nonnegative integer matrices satisfying the constraints:

$$\sum_{j=1}^{n-1} X_{ij} \le r_i \qquad \text{for all } i < m, \qquad \sum_{i=1}^{m-1} X_{ij} \le c_j \qquad \text{for all } j < n;$$

$$\sum_{i=1}^{m-1} \sum_{j=1}^{n-1} X_{ij} \ge \sum_{i=1}^{m-1} r_i - c_n.$$

For $X \in I(r, c)$, we will still refer to the quantities X_{in} and X_{mj}, but it will be with the understanding that they are defined in terms of the other entries, via

$$X_{in} \equiv r_i - \sum_{j=1}^{n-1} X_{ij} \qquad \text{for all } i < m, \qquad X_{mj} \equiv c_j - \sum_{i=1}^{m-1} X_{ij} \qquad \text{for all } j \le n.$$

Using this convention, the rounding method (which we described in Section 1) simply consists of rounding each of the $(m - 1)(n - 1)$ coordinates to the nearest integer. Furthermore, the convex set K' from which we perform continuous sampling will be a full-dimensional subset of $\mathbf{R}^{(m-1)(n-1)}$.

Recall that for integer matrices X and a convex set K', vol(X) is defined as the volume of points in K' which round to X. Thus, vol(X) is equal to the volume of the intersection of K' and the $(m - 1)(n - 1)$-dimensional unit hypercube centered at X. This leads to the following appealing characterization of vol(X). Let $\mathscr{E}$ be a random $(m - 1) \times (n - 1)$

matrix, whose entries are mutually independent and have the uniform distribution over $[-1/2, 1/2]$. Then for all integer matrices X we have $\mathrm{vol}(X) = \Pr(X + \mathscr{E} \in K')$. Thus, $\mathrm{vol}(X)$ is equal to the probability that, if we perturb X by adding a small random variable to each entry, then the result is in K'.

Now, in order for the output of our algorithm to have an almost uniform distribution, K' must be large enough so that $\mathrm{vol}(X)$ is nearly 1 for all $X \in I(r, c)$. In light of the above, this means that if we take any X in $I(r, c)$ and add $\mathscr{E}$, then the result must be in K' with high probability. We are now ready to define K'. Let $0 < \epsilon < 1/2$ be an error parameter, and let

$$C_1 = \frac{1}{2}\log\left(\frac{2}{\epsilon}\right), \qquad C_2 = \frac{\log(4/\epsilon)}{2\log m} + \frac{1}{2}, \qquad C_3 = \frac{\log(4/\epsilon)}{2\log n} + \frac{1}{2}.$$

Let K' be the set of real, $(m-1) \times (n-1)$ matrices Y satisfying

$$Y_{ij} \geq -1/2,$$
$$Y_{mn} \geq -\sqrt{C_1 mn},$$
$$Y_{in} \geq -\sqrt{C_2 n \log m}, \qquad Y_{mj} \geq -\sqrt{C_3 m \log n}, \tag{1}$$

for all $i < m$ and $j < n$. The reasons behind our choices for the above parameters should become clear after we prove the following lemma.

Lemma 1. *For any $X \in I(r, c)$, we have $1 \geq \mathrm{vol}(X) \geq 1 - \epsilon$.*

Proof. Let $X \in I(r, c)$ and let $X' = X + \mathscr{E}$. We want to show that $\Pr(X' \notin K')$ is at most ϵ. Now, $X'_{ij} \geq -1/2$ for all $i < m$ and $j < n$, since $|\mathscr{E}_{ij}| \leq 1/2$. We also have $X'_{in} = X_{in} - \sum_{j=i}^{n-1} \mathscr{E}_{ij} \geq -\sum_{j=i}^{n-1} \mathscr{E}_{ij}$ for all $i < m$. Recall Hoeffding's bounds [5]: Let $\{Y_j\}_{j=1}^{k}$ be independent, mean-zero random variables in $[-s, s]$. Then for all $A > 0$ we have $\Pr(\sum_{j=1}^{k} Y_j > A) \leq \exp(\frac{-A^2}{2s^2k})$. Applying Hoeffding's bounds to the $\mathscr{E}_{ij}$ gives

$$\Pr(X'_{in} < -\sqrt{C_2 n \log m}) \leq e^{-2C_2 \log m} = m^{-2C_2}, \qquad \text{for all } i < m.$$

Hence the probability that some X'_{in} is too small is at most $m^{1-2C_2} = \epsilon/4$. Applying Hoeffding's bounds again, this time to the column sums, gives

$$\Pr(X'_{mj} < -\sqrt{C_3 m \log n}) \leq e^{-2C_3 \log n} = n^{-2C_3}, \qquad \text{for all } j < n.$$

Hence the probability that some X'_{mj} is too small is also at most $\epsilon/4$. Finally, note that $X'_{mn} = X_{mn} + \sum_{i=1}^{m-1}\sum_{j=1}^{n-1} \mathscr{E}_{ij}$. Hence, Hoeffding's bounds imply that $\Pr(X'_{mn} < -\sqrt{C_1 mn}) \leq e^{-2C_1} = \epsilon/2$. Putting this all together, we conclude that $\Pr(X' \notin K') \leq \epsilon/4 + \epsilon/4 + \epsilon/2 = \epsilon$, so $\mathrm{vol}(X) \geq 1 - \epsilon$. ∎

We will call the algorithm that repeatedly samples from K' and rounds each sample to an integer point Algorithm A. Let Z be the random integer matrix produced by a single trial of Algorithm A. The preceding lemma implies that the conditional distribution of Z, given that it is in $I(r, c)$, is almost uniform. Thus, the final output of Algorithm A will indeed be an almost uniform sample from $I(r, c)$.

3. MAIN THEOREM

In order to bound the running time of the algorithm, we must bound the expected number of trials taken before the random sample Z is in $I(r, c)$. We now state the main result of this paper.

Theorem 2. *Suppose that the row and column sums satisfy $r_i = \Omega(n^{3/2} m \log m)$ and $c_j = \Omega(m^{3/2} n \log n)$ for all i and j. Then the expected number of trials before algorithm A generates a sample point $Z \in I(r, c)$ is $O(1/\epsilon)$.*

Proof. Let $D \in (0, 1]$ and suppose that the row and column sums satisfy $r_i \geq Dn^{3/2}m \log m$ and $c_j \geq Dm^{3/2}n \log n$ for all i and j. We will show that the expected number of trials is $\exp(O(1/D^2)) \times 1/\epsilon$.

Algorithm A repeatedly generates random integer matrices Z until one of them is in $I(r, c)$. Thus, the number of trials is a geometric random variable with parameter $\Pr(Z \in I(r, c))$. Let $\bar{I}(r, c)$ denote the set of integer matrices X having $\mathrm{vol}(X) > 0$, i.e., the points which have some $Y \in K'$ rounding to them. Then the expected number of trials $1/\Pr(Z \in I(r,c))$ is equal to

$$\frac{\mathrm{vol}(K')}{\mathrm{vol}(I(r, c))} = \frac{\mathrm{vol}(\bar{I}(r, c))}{\mathrm{vol}(I(r, c))}.$$

Suppose that $X \in \bar{I}(r, c)$. Then, by (1), X must satisfy $X_{ij} \geq 0$ for all $i < m$ and $j < n$. Thus, X is also in $I(r, c)$ if and only if it satisfies $X_{in} \geq 0$ for all $1 \leq i \leq m$ and $X_{mj} \geq 0$ for all $1 \leq j \leq n$. Thus X, when thought of as an $m \times n$ matrix, is in $I(r, c)$ when it has no negative entry anywhere in its last row or column. We must show that such points form a nonnegligible fraction of $\bar{I}(r, c)$.

A sketch of our argument is as follows. First, consider the random variable Z_{1n}. Since the row and colum sums are large, the probabilities $\Pr(Z_{1n} = k)$ will remain roughly constant over a long interval in k. Thus, since the number of possible negative values for Z_{1n} is limited to about n, the probability that Z_{1n} will take a nonnegative value is quite large. Of course, a similar argument will hold for the other Z_{in} and Z_{mj}, so Z will stand a good chance of being in $I(r, c)$.

Instead of working directly with the vol function, we will find it easier to work with an upper bound on vol which is based on Hoeffding's bounds. As in the proof of Lemma 1, we will use the fact that $\mathrm{vol}(X) = \Pr(X' \in K')$, where $X' = X + \mathscr{E}$ and $\mathscr{E}$ is a random matrix with uniform $[-1/2, 1/2]$ entries. For real numbers x, let x^+ and x^- denote the positive and negative parts of x, respectively:

$$x^+ = \max(0, x), \qquad x^- = \max(0, -x).$$

Note that the random variables $\{X'_{in}\}_{i=1}^{m-1}$ are independent, and so are $\{X'_{mj}\}_{j=1}^{n-1}$. Thus Hoeffding's bounds give the following three upper bounds on $\mathrm{vol}(X)$:

$$\mathrm{vol}(X) \le \prod_{i=1}^{m-1} \Pr(X'_{in} \ge -C'_2) \le \exp\left(\frac{-2\sum_{i<m}((X_{in}+C'_2)^-)^2}{n}\right),$$

$$\mathrm{vol}(X) \le \prod_{j=1}^{n-1} \Pr(X'_{mj} \ge -C'_3) \le \exp\left(\frac{-2\sum_{j<n}((X_{mj}+C'_3)^-)^2}{m}\right),$$

$$\mathrm{vol}(X) \le \Pr(X'_{mn} \ge -C'_1) \le \exp\left(\frac{-2\,((X_{mn}+C'_1)^-)^2}{mn}\right),$$

where $C'_2 = \sqrt{C_2 n \log m}$, $C'_3 = \sqrt{C_3 m \log n}$, and $C'_1 = \sqrt{C_1 mn}$. Putting these bounds together (taking the geometric mean of the three bounds), we get $\mathrm{vol}(X) \le w(X)$, where

$$w(X) = \exp\left\{-\frac{2}{3}\left[\frac{\sum_{i<m}((X_{in}+C'_2)^-)^2}{n} + \frac{\sum_{j<n}((X_{mj}+C'_3)^-)^2}{m} + \frac{((X_{mn}+C'_1)^-)^2}{mn}\right]\right\}. \quad (2)$$

We will call w a *weight* function. Note that for all integer points X we have $0 \le \mathrm{vol}(X) \le w(X) \le 1$, and Lemma 1 implies that $\mathrm{vol}(X) > 1/2$ for $X \in I(r, c)$ (recall that $\epsilon < 1/2$). Hence $w(I(r, c)) < 2 \times \mathrm{vol}(I(r, c))$. It follows that the expected number of trials $\frac{\mathrm{vol}(\bar{I}(r,c))}{\mathrm{vol}(I(r,c))} < 2 \times \frac{w(\bar{I}(r,c))}{w(I(r,c))}$. Thus, our task reduces to giving an upper bound on the quantity $\frac{w(\bar{I}(r,c))}{w(I(r,c))}$.

Recall that $\bar{I}(r, c) - I(r, c)$ consists of the points in $\bar{I}(r, c)$ which have a negative entry somewhere in their last row or column. Let W_0 denote the set of matrices X in $\bar{I}(r, c)$ which have $X_{mn} \ge 0$. For all i in $\{1, \ldots, m-1\}$, define $W_i = \{X \in W_{i-1} : X_{in} \ge 0\}$. Finally, for all j in $\{m, m+n-2\}$, define $W_j = \{X \in W_{j-1} : X_{m,j-m+1} \ge 0\}$. Then $\bar{I}(r, c) \supset W_0 \supset W_1 \cdots \supset W_{m+n-2} = I(r, c)$. Our strategy will be to write $\frac{w(\bar{I}(r,c))}{w(I(r,c))}$ as the product

$$\left(\frac{w(\bar{I}(r, c))}{w(W_0)}\right) \times \left(\frac{w(W_0)}{w(W_1)}\right) \times \left(\frac{w(W_{m+n-3})}{w(W_{m+n-2})}\right) \quad (3)$$

and then show that each factor is not too large.

For a matrix X, let X^- denote the matrix whose entries are the negative parts of the entries of X. To bound each factor in (3), we will use the following lemma.

Lemma 3. *Let I be a set of integer $m \times n$ matrices which is closed upward in the sense that if $X^- \le Y^-$ and $Y \in I$, then $X \in I$. Fix integers $k \le m$ and $l \le n$, positive real numbers C, and $\alpha \ge 1$, and suppose that $\nu : I \to \mathbf{R}^+$ is a weight function with the property that if $X^- \le Y^-$ and $Y_{kl} = X_{kl} - 1$; then $\frac{\nu(Y)}{\nu(X)} \le \exp\left(\frac{-2(X_{kl}+C)^-}{\alpha^2}\right)$. Let β be a positive real number and suppose that for every $X \in I$ we have $\sum_j X_{kj} \ge 2\alpha\beta n$ and $\sum_i X_{il} \ge 2\alpha\beta m$. Define $\bar{I} = \{X \in I : X_{kl} \ge 0\}$. Then*

$$\frac{\nu(I)}{\nu(\bar{I})} \le \exp\left(\frac{3[C + \alpha\max(3, 1 + 2/\beta)]}{\alpha\beta}\right).$$

Proof. W.l.o.g., suppose that $k = m$ and $l = n$. For integers s, let $V_s = \{X \in I : X_{mn} = s\}$. Then we want to bound

$$\frac{\sum_s \nu(V_s)}{\sum_{s\geq 0} \nu(V_s)}. \tag{4}$$

We will do this by giving an upper bound on $\frac{\nu(V_{s-1})}{\nu(V_s)}$ for each s. What we want to show is that $\nu(V_{s-1}) \leq \alpha\nu(V_s)$, for a value of α which is as small as possible. We will do this in a way that is reminiscent of the standard technique in combinatorics in which one shows that two sets S_1 and S_2 satisfy $|S_1| \leq k|S_2|$ by giving a k-to-1 function from S_1 to S_2. What we will do here is give a *random* function (i.e., a function which is itself a random variable) from V_{s-1} to V_s with the property that for every $X \in V_s$, the *expected value* of the total weight of points mapping to X is at most $\alpha \times \nu(X)$.

For positive integers $a < m$ and $b < n$, let $T_{ab}(\cdot)$ be the transformation, acting on a matrix X, that increases X_{ab} and X_{mn} by 1 and reduces X_{an} and X_{mb} by 1. Let $M = \alpha\beta$, and for integers s with $s \leq M$, let f_s be a random function from V_{s-1} to V_s such that

$$f_s(X) = T_{ab}(X) \qquad \text{with probability } \frac{X_{an}^{+}X_{mb}^{+}}{(\sum_{i<m} X_{in}^{+})(\sum_{j<n} X_{mj}^{+})},$$

for all $a < m$ and $b < n$. Note that in the definition of f_s the random a and b which occur with nonzero probability will always satisfy $X_{an} > 0$ and $X_{mb} > 0$, so f_s will not increase the negative part of any entry. Hence if $Y = T_{ab}^{-1}(X)$ for some a and b (so that $Y_{mn} = s - 1$), the assumptions of the lemma imply that

$$\frac{\nu(Y)}{\nu(X)} \leq p(s),$$

where we define $p(s) = \exp\left\{\frac{-2(s+C)^{-}}{\alpha^2}\right\}$.

Thus for all $s \leq M$ and $X \in V_s$ we have

$$\begin{aligned}
\frac{\mathrm{E}(\nu(f_s^{-1}(X)))}{\nu(X)} &\leq \sum_{a=1}^{m-1}\sum_{b=1}^{n-1} \frac{(X_{an}^{+}+1)(X_{mb}^{+}+1)}{(\sum_{i<m} X_{in}^{+}+1)(\sum_{j<n} X_{mj}^{+}+1)} \times p(s) \\
&= \frac{[\sum_{a<m} X_{an}^{+} + m][\sum_{b<n} X_{mb}^{+} + n]}{[\sum_{a<m} X_{an}^{+} + 1][\sum_{b<n} X_{mb}^{+} + 1]} \times p(s) \\
&\leq \left(1 + \frac{1}{\alpha\beta}\right)^2 p(s) \\
&\leq \exp\left(\frac{2}{\alpha\beta}\right) p(s),
\end{aligned}$$

where the second inequality follows from the facts that $[\sum_{a=1}^{m-1} X_{an}^{+}] \geq c_n - X_{mn} \geq \alpha\beta m$ and $[\sum_{b=1}^{n-1} X_{mb}^{+}] \geq r_m - X_{mn} \geq \alpha\beta n$ (recall that $X_{mn} \leq M = \alpha\beta$). It follows that for all $s \leq M$ we have

$$\frac{\nu(V_{s-1})}{\nu(V_s)} \leq \exp\left(\frac{2}{\alpha\beta}\right) p(s).$$

Let $\beta' = \min(\beta, 1)$ and let $A = 2/\beta'$. Then for $s \le -C - A\alpha$ we have

$$\frac{\nu(V_{s-1})}{\nu(V_s)} \le \exp\left(\frac{2}{\alpha\beta} - \frac{2A}{\alpha}\right) = \exp\left(\frac{2}{\alpha\beta} - \frac{4}{\alpha\beta'}\right) \le \exp\left(-\frac{2}{\alpha}\right).$$

Thus for all s we have

$$\frac{\nu(V_{s-1})}{\nu(V_s)} \le \begin{cases} \exp\left(-\dfrac{2}{\alpha}\right) & \text{if } s \le -C - A\alpha; \\ \exp\left(\dfrac{2}{\alpha\beta}\right) & \text{otherwise.} \end{cases} \tag{5}$$

Note that the quantity (4) we are trying to bound is an increasing function of the ratios $\frac{\nu(V_{s-1})}{\nu(V_s)}$. Thus we can assume that the inequalities in (5) are actually equalities, since this will only increase (4). We obtain

$$\frac{\sum_s \nu(V_s)}{\sum_{s\ge 0} \nu(V_s)} \le \frac{\sum_{j\ge 0} \exp(-2j/\alpha) + \sum_{j=1}^{M+C+A\alpha} f(j)}{\sum_{j=C+A\alpha}^{M+C+A\alpha} f(j)}, \tag{6}$$

where $f(j) = \exp(\frac{-2j}{\alpha\beta})$. In the sums, the index j represents the distance between s and $-C - A\alpha$. Since the first sum in the numerator is less than α, the RHS of (6) is less than

$$\frac{\sum_{j=-\alpha}^{M+C+A\alpha} f(j)}{\sum_{j=C+A\alpha}^{M+C+A\alpha} f(j)} = \frac{1 - f(M + \Delta + 1)}{1 - f(M+1)} f(-\Delta),$$

where $\Delta = C + (A + 1)\alpha$. Note that

$$\begin{aligned}
\frac{1 - f(M + \Delta + 1)}{1 - f(M + 1)} &= 1 + \frac{f(M+1) - f(M+1+\Delta)}{1 - f(M+1)} \\
&\le 1 + 2[f(M+1) - f(M+1+\Delta)] \\
&\le 1 + 2|f'(M+1)|\Delta \\
&\le 1 + 2 \times \left[\frac{2}{\alpha\beta} \times f(M+1)\right] \times \Delta \\
&\le 1 + \frac{\Delta}{\alpha\beta} \\
&\le \exp\left(\frac{\Delta}{\alpha\beta}\right) \\
&= f\left(-\frac{1}{2}\Delta\right),
\end{aligned}$$

where the first and fourth inequalities follow from $f(M + 1) \le e^{-2}$, and the second inequality is Taylor's theorem. Thus

$$\frac{\sum_s \nu(V_s)}{\sum_{s\geq 0} \nu(V_s)} \leq f\left(-\frac{1}{2}\Delta\right)f(-\Delta) = \exp\left(\frac{3[C + \alpha \max(3, 1 + 2/\beta)]}{\alpha\beta}\right). \tag{7}$$

■

Now we will use Lemma 3 to bound the first factor in (3), namely $\frac{w(\bar{I}(r,c))}{w(W_0)}$. Note that if $X^- \leq Y^-$ and $Y_{mn} = X_{mn} - 1$, then

$$\frac{w(Y)}{w(X)} = \exp\left\{\frac{((X_{mn} + C_1')^-)^2 - ((Y_{mn} + C_1')^-)^2}{\frac{3}{2}mn}\right\} \leq \exp\left\{\frac{-2(X_{mn} + C_1')^-}{\frac{3}{2}mn}\right\}$$

Thus the conditions of Lemma 3 are satisfied (recall $m \leq n$), with

$$\alpha = \sqrt{\frac{3}{2}mn}, \qquad \beta = \frac{D}{\sqrt{6}}\sqrt{m}\log m, \qquad C = C_1' = \sqrt{C_1 mn},$$

$$I = \bar{I}(r, c), \qquad \nu = w.$$

Thus the lemma implies that

$$\frac{w(W_0)}{w(\bar{I}(r, c))} \leq \exp\left(\frac{3\left[\sqrt{C_1 mn} + \sqrt{\frac{3}{2}mn} \times \max\left(3, 1 + \frac{2\sqrt{6}}{D\sqrt{m}\log m}\right)\right]}{\frac{1}{2}Dm\sqrt{n}\log m}\right).$$

Next we will bound the second ratio in (3), namely $\frac{w(W_0)}{w(W_1)}$. Note that if $X^- \leq Y^-$ and $Y_{1n} = X_{1n} - 1$, then $\frac{w(Y)}{w(X)} \leq \exp\left\{\frac{-2(X_{1n} + C_2')^-}{\frac{3}{2}n}n\right\}$. Thus the conditions of Lemma 3 are satisfied, with

$$\alpha = \sqrt{\frac{3}{2}n}, \qquad \beta = \frac{D}{\sqrt{6}}m\log m, \qquad C = C_2' = \sqrt{C_2 n\log m},$$

$$I = W_0, \qquad \nu = w_0,$$

where w_0 is the restriction of w to W_0. Thus the lemma implies that

$$\frac{w(W_0)}{w(W_1)} \leq \exp\left(\frac{3\left[\sqrt{C_2 n\log m} + \sqrt{\frac{3}{2}n} \times \max\left(3, 1 + \frac{2\sqrt{6}}{Dm\log m}\right)\right]}{\frac{1}{2}Dm\sqrt{n}\log m}\right).$$

A similar argument gives the same bound for $\frac{w(W_i)}{w(W_{i+1})}$ for $i = 1, \ldots, m - 2$.

Next we will bound $\frac{w(W_{m-1})}{w(W_m)}$. Note that if $X^- \leq Y^-$ and $Y_{m1} = X_{m1} - 1$, then $\frac{w(Y)}{w(X)} \leq$

$\exp\left\{\frac{-2(X_{m1}+C_3')^-}{\frac{3}{2}m}\right\}$. Also, since r_m is the largest row sum, we have $r_m \geq \frac{1}{m}\sum_{i=1}^m r_i = \frac{1}{m}\sum_{j=1}^n c_j \geq D\sqrt{m}\, n^2\log n$. Thus, the conditions of Lemma 3 are satisfied, with

$$\alpha = \sqrt{\frac{3}{2}m}, \qquad \beta = \frac{D}{\sqrt{6}}\, n \log n, \qquad C = C_3' = \sqrt{C_3 m \log n},$$

$$I = W_{m-1}, \qquad \nu = w_{m-1},$$

where w_{m-1} is the restriction of w to W_{m-1}. Thus the lemma implies that

$$\frac{w(W_{m-1})}{w(W_m)} \leq \exp\left(\frac{3\left[\sqrt{C_3 m \log n} + \sqrt{\frac{3}{2}m} \times \max\left(3,\, 1 + \frac{2\sqrt{6}}{Dn\log n}\right)\right]}{\frac{1}{2}Dn\sqrt{m}\log n}\right).$$

A similar argument gives the same bound for $\frac{w(W_j)}{w(W_{j+1})}$ for $j = m, \ldots, m+n-3$.

After plugging all of our bounds into (3), some easy calculations (using the facts that $m, n \geq 2$, and $C_1, C_2, C_3 \leq \log(4/\epsilon)$) give

$$\frac{w(\bar{I}(r, c))}{w(I(r, c))} \leq \exp\left(\frac{A}{D^2}\right) \times \exp\left(\frac{B}{D}\sqrt{\log\left(\frac{1}{\epsilon}\right)}\right),$$

for some constants A and B. Let $\gamma = \frac{B}{D\sqrt{\log(1/\epsilon)}}$. Then

$$\exp\left(\frac{B}{D}\sqrt{\log\left(\frac{1}{\epsilon}\right)}\right) = \frac{1}{\epsilon} \times \exp\left(\frac{B^2}{D^2} \times \frac{\gamma - 1}{\gamma^2}\right)$$
$$\leq \frac{1}{\epsilon} \times \exp\left(\frac{B^2}{4D^2}\right),$$

where the inequality holds because $\frac{\gamma-1}{\gamma^2} \leq \frac{1}{4}$ for all real numbers γ. Hence

$$\frac{w(\bar{I}(r, c))}{w(I(r, c))} \leq \frac{1}{\epsilon} \times \exp\left(\frac{A + \frac{B^2}{4}}{D^2}\right) = \frac{1}{\epsilon} \times \exp\left(O\left(\frac{1}{D^2}\right)\right).$$

■

Under the assumptions of the theorem, the (expected) running time of Algorithm A is of the form

$$q(m, n) \times O(1/\epsilon),$$

where $q(m, n)$ is the (polynomial) running time of a single trial. Hence the algorithm runs in time which is a polynomial function of m, n, and $1/\epsilon$.

Note that we do not have the logarithmic dependence on $1/\epsilon$ that can be achieved in many sampling algorithms that use direct applications of Markov Chains. However, in most applications of random sampling (e.g., [11, 6]), logarithmic dependence on $1/\epsilon$ is not necessary.

If the continuous samples from the convex set K' were *perfectly* uniform, then by Lemma 1 the parameter ϵ would be an upper bound on the total variation distance between the sample produced by Algorithm A and the uniform distribution over $I(r, c)$. However, in all general algorithms for sampling continuously from convex set, the samples are only *almost* uniform. Algorithm A will therefore only guarantee that the distance from uniform is at most $\epsilon + \delta$ for some small δ.

ACKNOWLEDGMENTS

I would like to thank Alistair Sinclair for telling me about the contingency tables problem and for many helpful conversations. I would also like to thank the referees for their thoroughness and for their excellent suggestions.

REFERENCES

[1] P. Diaconis and B. Efron, Testing for independence in a two-way table: new interpretations of the chi-squared statistic, Ann Stat 13 (1985), 845–913.

[2] M. E. Dyer, A. M. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, J ACM 38 (1991), 1–17.

[3] M. Dyer and C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables, Theoret Comput Sci 246 (2000), 265–278.

[4] M. E. Dyer, R. Kannan, and J. Mount, Sampling contingency tables, Random Struct Alg 10 (1997), 487–506.

[5] W. Hoeffding, Probability inequalities for sums of bounded random variables, J Am Stat Assoc 58 (1963), 13–30.

[6] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theoret Comput Sci 43 (1986), 169–188.

[7] R. Kannan, L. Lovász, and M. Simonovits, Random walks and an $O^*(n^5)$ volume algorithm for convex bodies, Random Struct Alg 11 (1997), 1–50.

[8] R. Kannan and S. Vempala, Sampling lattice points, Proc ACM Symp Theory Comput, 1997, 696–700.

[9] L. Lovász, Hit and run mixes fast, Math Program 86 (1999), 443–461.

[10] L. Lovász and R. Kannan, Faster mixing via average conductance, Proc ACM Symp Theory Comput, 1999, 282–287.

[11] A. J. Sinclair, Randomised algorithms for counting and generating combinatorial structures, Advances in Theoretical Computer Science, Birkhäuser, Boston, 1993.