ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Advances in Mathematics 224 (2010) 316–339
ADVANCES IN Mathematics
www.elsevier.com/locate/aim

On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries ☆

Alexander Barvinok

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043, USA

Received 29 July 2008; accepted 1 December 2009

Available online 14 December 2009

Communicated by Gil Kalai

Abstract

We consider the set $\Sigma(R, C)$ of all $m \times n$ matrices having 0–1 entries and prescribed row sums $R = (r_1, \ldots, r_m)$ and column sums $C = (c_1, \ldots, c_n)$. We prove an asymptotic estimate for the cardinality $|\Sigma(R, C)|$ via the solution to a convex optimization problem. We show that if $\Sigma(R, C)$ is sufficiently large, then a random matrix $D \in \Sigma(R, C)$ sampled from the uniform probability measure in $\Sigma(R, C)$ with high probability is close to a particular matrix $Z = Z(R, C)$ that maximizes the sum of entropies of entries among all matrices with row sums R, column sums C and entries between 0 and 1. Similar results are obtained for 0–1 matrices with prescribed row and column sums and assigned zeros in some positions.

Keywords: 0–1 matrices; Entropy; Asymptotic estimates; Integer flows

1. Introduction and main results

Matrices with 0–1 entries and prescribed row and column sums is a classical object which appears in many branches of pure and applied mathematics. In combinatorics, such matrices encode hypergraphs with prescribed degrees of vertices and related structures, see, for example, [25]. In algebra, certain structural constants in the ring of symmetric functions and, consequently, in the representation theory of the symmetric and general linear groups are expressed as numbers of 0–1

☆ This research was partially supported by NSF Grants DMS 0400617 and DMS 0856640 and a United States–Israel BSF Grant 2006377.

E-mail address: barvinok@umich.edu.

0001-8708/$ – see front matter
doi:10.1016/j.aim.2009.12.001

matrices with prescribed row and column sums, see [21, Chapter 1]. In statistics, 0–1 matrices with prescribed row and column sums are known as *binary contingency tables*, see [9].

Let $R = (r_1, \ldots, r_m)$ be a positive integer m-vector and let $C = (c_1, \ldots, c_n)$ be a positive integer n-vector such that

$$\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = N \quad \text{and}$$
$$0 < r_i < n \quad \text{for } i = 1, \ldots, m \quad \text{and} \quad 0 < c_j < m \quad \text{for } j = 1, \ldots, n.$$

Let $\Sigma(R, C)$ be the set of all $m \times n$ matrices (binary contingency tables) $D = (d_{ij})$ such that

$$\sum_{j=1}^{n} d_{ij} = r_i \quad \text{for } i = 1, \ldots, m, \qquad \sum_{i=1}^{m} d_{ij} = c_j \quad \text{for } j = 1, \ldots, n$$
$$\text{and} \quad d_{ij} \in \{0, 1\}.$$

In words: $\Sigma(R, C)$ is the set of 0–1 matrices with row sums R and column sums C. Vectors R and C are called *margins* of a matrix $D \in \Sigma(R, C)$.

Our first main result provides an estimate of the cardinality of $\Sigma(R, C)$.

Theorem 1. *Let us define the function*

$$F(\mathbf{x}, \mathbf{y}) = \left(\prod_{i=1}^{m} x_i^{-r_i}\right)\left(\prod_{j=1}^{n} y_j^{-c_j}\right)\left(\prod_{ij} (1 + x_i y_j)\right)$$
$$\textit{for } \mathbf{x} = (x_1, \ldots, x_m) \textit{ and } \mathbf{y} = (y_1, \ldots, y_n)$$

and let

$$\alpha(R, C) = \inf_{\substack{x_1, \ldots, x_m > 0 \\ y_1, \ldots, y_n > 0}} F(\mathbf{x}, \mathbf{y}).$$

Then for the number $|\Sigma(R, C)|$ of $m \times n$ zero–one matrices with row sums R and column sums C we have

$$\alpha(R, C) \geqslant \left|\Sigma(R, C)\right| \geqslant \frac{(mn)!}{(mn)^{mn}} \left(\prod_{i=1}^{m} \frac{(n - r_i)^{n - r_i}}{(n - r_i)!}\right)\left(\prod_{j=1}^{n} \frac{c_j^{c_j}}{c_j!}\right) \alpha(R, C).$$

Let us estimate the ratio between the lower and the upper bounds for $|\Sigma(R, C)|$ using Stirling's formula

$$s! s^{-s} = e^{-s} \sqrt{2\pi s}\bigl(1 + O(s^{-1})\bigr).$$

Since

$$e^{-mn}\left(\prod_{i=1}^{m} e^{n-r_i}\right)\left(\prod_{j=1}^{n} e^{c_j}\right)=1,$$

the "e^{-s}" contributions from Stirling's formula cancel each other out and we obtain

$$\alpha(R,C) \geqslant \left|\Sigma(R,C)\right| \geqslant (mn)^{-\gamma(m+n)}\alpha(R,C)$$

for some absolute constant $\gamma > 0$.

We note that in many interesting cases we have $|\Sigma(R,C)| = 2^{\Omega(mn)}$, see also Section 3.1, in which case the estimate of Theorem 1 captures the logarithmic order of $|\Sigma(R,C)|$.

Let us substitute $x_i = e^{s_i}$, $y = e^{t_i}$ in $F(\mathbf{x},\mathbf{y})$. Then $\ln F(\mathbf{x},\mathbf{y}) = G(\mathbf{s},\mathbf{t})$, where

$$G(\mathbf{s},\mathbf{t}) = -\sum_{i=1}^{m} r_i s_i - \sum_{j=1}^{n} c_j t_j + \sum_{ij} \ln\left(1+e^{s_i+t_j}\right)$$
$$\text{for } \mathbf{s}=(s_1,\dots,s_m) \text{ and } \mathbf{t}=(t_1,\dots,t_n).$$

One can observe that $G(\mathbf{s},\mathbf{t})$ is a convex function on $\mathbb{R}^m \times \mathbb{R}^n$, hence to compute the infimum of $G(\mathbf{s},\mathbf{t})$ one can use any of the efficient convex optimization algorithms, see, for example, [23].

Suppose that margins R, C are such that the set $\Sigma(R,C)$ is not empty and let us consider $\Sigma(R,C)$ as a finite probability space with the uniform measure. Let us pick a random matrix $D \in \Sigma(R,C)$. What is D likely to look like? This question is of some interest to statistics: a binary contingency table $D = (d_{ij})$ may represent certain statistical data (for example, d_{ij} may be equal to 1 or 0 depending on whether or not Darwin finches of the i-th species can be found on the j-th Galapagos island, as in [9]). One can condition on the row and column sums and ask what is special about a particular table $D \in \Sigma(R,C)$, considering all tables in $\Sigma(R,C)$ as equiprobable, see [9]. To answer this question we need to know what a random table $D \in \Sigma(R,C)$ looks like.

We prove that with high probability D is close to a particular matrix Z with row sums R and column sums C and entries between 0 and 1, which we call the *maximum entropy matrix*.

1.1. The maximum entropy matrix

For $0 \leqslant x \leqslant 1$ let us consider the *entropy function*

$$H(x) = x\ln\frac{1}{x} + (1-x)\ln\frac{1}{1-x}.$$

As is known, H is a strictly concave function with $H(0) = H(1) = 0$.

For an $m \times n$ matrix $X = (x_{ij})$ such that $0 \leqslant x_{ij} \leqslant 1$ for all i, j, we define

$$H(X) = \sum_{ij} H(x_{ij}).$$

Assume that $\Sigma(R,C)$ is non-empty. Let us consider the polytope $\mathcal{P}(R,C)$ of matrices $X = (x_{ij})$ such that

$$\sum_{j=1}^{n} x_{ij} = r_i \quad \text{for } i = 1, \dots, m, \qquad \sum_{i=1}^{m} x_{ij} = c_j \quad \text{for } j = 1, \dots, n$$
$$\text{and} \quad 0 \leqslant x_{ij} \leqslant 1 \quad \text{for all } i, j.$$

Since $H(X)$ is strictly concave, it attains a unique maximum $Z = Z(R, C)$ on $\mathcal{P}(R, C)$, which we call the *maximum entropy matrix* with margins (R, C).

For example, if all r_i are equal, then by the symmetry argument we must have $Z = (z_{ij})$ where $z_{ij} = c_j/m$ for all i, j.

The following observation characterizes the maximum entropy matrix as the solution to the problem that is convex dual to the optimization problem of Theorem 1.

Lemma 2. *Suppose that the polytope $\mathcal{P}(R, C)$ has a non-empty interior, that is, contains a matrix $Y = (y_{ij})$ such that $0 < y_{ij} < 1$ for all i, j. Then the infimum $\alpha(R, C)$ in Theorem 1 is attained at a particular point $\mathbf{x}^* = (\xi_1, \dots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \dots, \eta_n)$. For the maximum entropy matrix $Z = (z_{ij})$ we have*

$$z_{ij} = \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} \quad \textit{for all } i, j \tag{1}$$

and, moreover,

$$\alpha(R, C) = e^{H(Z)}. \tag{2}$$

Conversely, if the infimum $\alpha(R, C)$ in Theorem 1 is attained at a certain point $\mathbf{x}^* = (\xi_1, \dots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \dots, \eta_n)$ then for the maximum entropy matrix $Z = (z_{ij})$ Eqs. (1) and (2) hold.

The condition that the polytope $\mathcal{P}(R, C)$ has a non-empty interior is equivalent to the requirement that for every choice of $1 \leqslant k \leqslant m$ and $1 \leqslant l \leqslant n$ there is a matrix $D^0 \in \Sigma(R, C)$, $D^0 = (d^0_{ij})$, such that $d^0_{kl} = 0$ and there is a matrix $D^1 \in \Sigma(R, C)$, $D^1 = (d^1_{ij})$, such that $d^1_{kl} = 1$. One can take Y to be the average of all matrices $D \in \Sigma(R, C)$. In other words, we require the set $\Sigma(R, C)$ to be reasonably large. We also observe that if $r_i c_j < N$ for all i, j (recall that N is the total sum of the matrix entries) one can choose $y_{ij} = r_i c_j / N$.

We prove that with high probability a random matrix $D \in \Sigma(R, C)$ is close to the maximum entropy matrix Z as far as sums over subsets of entries are concerned.

For a subset

$$S \subset \{(i, j): \ i = 1, \dots, m, \ j = 1, \dots, n\}$$

and an $m \times n$ matrix $A = (a_{ij})$, let us denote

$$\sigma_S(A) = \sum_{(i,j) \in S} a_{ij},$$

the sum of the entries of A indexed by S.

In what follows, we are interested in the case of the density N/mn separated from 0. Without loss of generality, we assume that $n \geqslant m$.

Theorem 3. *Let us fix numbers $\kappa > 0$ and $0 < \delta < 1$. Then there exists a number $q = q(\kappa, \delta)$ such that the following holds.*

Let (R, C) be margins such that $n \geqslant m > q$ and the polytope $\mathcal{P}(R, C)$ has a non-empty interior, and let $Z \in \mathcal{P}(R, C)$ be the maximum entropy matrix. Let $S \subset \{(i, j)\colon i = 1, \dots, m,\ j = 1, \dots, n\}$ be a subset such that $\sigma_S(Z) \geqslant \delta mn$ and let

$$\epsilon = \delta \frac{\ln n}{\sqrt{m}}.$$

If $\epsilon \leqslant 1$ then

$$\mathbf{Pr}\bigl\{D \in \Sigma(R, C)\colon\ (1 - \epsilon)\sigma_S(Z) \leqslant \sigma_S(D) \leqslant (1 + \epsilon)\sigma_S(Z)\bigr\} \geqslant 1 - 2n^{-\kappa n}.$$

Let us associate with a non-negative, non-zero $m \times n$ matrix $A = (a_{ij})$ a finite probability space on the ground set $\{(i, j)\colon\ i = 1, \dots, m,\ j = 1, \dots, n\}$ with $\mathbf{Pr}\{(i, j)\} = a_{ij}/N$, where $N > 0$ is the total sum of matrix entries. Theorem 3 asserts that the probability space associated with the maximum entropy matrix Z reasonably well approximates the probability space associated with a random binary contingency table $D \in \Sigma(R, C)$ as far as events S whose probability is separated from 0 are concerned.

The following interpretation of the maximum entropy matrix was suggested to the author by J.A. Hartigan, see [4].

Theorem 4. *Let $Z = (z_{ij})$ be the $m \times n$ maximum entropy matrix with margins (R, C) and let us suppose that the polytope $\mathcal{P}(R, C)$ has a non-empty interior. Let $X = (x_{ij})$ be the random $m \times n$ matrix of independent Bernoulli random variables such that*

$$\mathbf{E}X = Z.$$

In other words, $\mathbf{Pr}\{x_{ij} = 1\} = z_{ij}$ and $\mathbf{Pr}\{x_{ij} = 0\} = 1 - z_{ij}$ independently for all i, j. Then the probability mass function of X is constant on the set $\Sigma(R, C)$ of binary contingency tables with margins (R, C), and, moreover,

$$\mathbf{Pr}\{X = D\} = e^{-H(Z)} \quad \text{for all } D \in \Sigma(R, C).$$

2. Extensions and ramifications

Our results hold in a somewhat greater generality. Let us fix an $m \times n$ non-negative matrix $W = (w_{ij})$, which we call the matrix of *weights*. Let us consider the following *partition function*

$$\bigl|\Sigma(R, C; W)\bigr| = \sum_{\substack{D \in \Sigma(R, C) \\ D = (d_{ij})}} \prod_{\substack{i, j \\ d_{ij} = 1}} w_{ij}.$$

In particular, if $w_{ij} = 1$ for all i, j then $|\Sigma(R, C; W)| = |\Sigma(R, C)|$. If $w_{ij} \in \{0, 1\}$ then the partition function counts binary contingency tables with zeros assigned to some positions: the value of $|\Sigma(R, C; W)|$ is equal to the number of $m \times n$ matrices $D = (d_{ij})$ such that the row sums of D are R, the column sums of D are C, $d_{ij} \in \{0, 1\}$ for all i, j, and, additionally, $d_{ij} = 0$ if

$w_{ij} = 0$. In combinatorial terms, the set $\Sigma(R, C; W)$ can be interpreted as the set of all subgraphs with prescribed degrees of vertices of a given bipartite graph. Binary contingency tables with preassigned zeros are of interest in statistics, see [9].

We prove the following result.

Theorem 5. *Let us define the function*

$$F(\mathbf{x}, \mathbf{y}; W) = \left(\prod_{i=1}^{m} x_i^{-r_i}\right)\left(\prod_{j=1}^{n} y_j^{-c_j}\right)\left(\prod_{ij}(1 + w_{ij} x_i y_j)\right)$$
for $\mathbf{x} = (x_1, \ldots, x_m)$ *and* $\mathbf{y} = (y_1, \ldots, y_n)$

and let

$$\alpha(R, C; W) = \inf_{\substack{x_1, \ldots, x_m > 0 \\ y_1, \ldots, y_n > 0}} F(\mathbf{x}, \mathbf{y}; W).$$

Then for the partition function $|\Sigma(R, C; W)|$ *we have*

$$\alpha(R, C; W) \geqslant \left|\Sigma(R, C; W)\right| \geqslant \frac{(mn)!}{(mn)^{mn}}\left(\prod_{i=1}^{m} \frac{(n - r_i)^{n - r_i}}{(n - r_i)!}\right)\left(\prod_{j=1}^{n} \frac{c_j^{c_j}}{c_j!}\right)\alpha(R, C; W).$$

As before, the function obtained as the result of the substitution $x_i = e^{t_i}$, $y_j = e^{s_j}$ in $\ln F(\mathbf{x}, \mathbf{y}; W)$,

$$G(\mathbf{s}, \mathbf{t}; W) = -\sum_{i=1}^{m} r_i s_i - \sum_{j=1}^{n} c_j t_j + \sum_{ij} \ln\left(1 + w_{ij} e^{s_i + t_j}\right)$$
for $\mathbf{s} = (s_1, \ldots, s_m)$ and $\mathbf{t} = (t_1, \ldots, t_n)$

is convex on $\mathbb{R}^m \times \mathbb{R}^n$, hence computing $\alpha(R, C; W)$ is a convex optimization problem.

Let us assume now that $w_{ij} \in \{0, 1\}$ for all (i, j) and let us consider the set $\Sigma(R, C; W)$ of all $m \times n$ binary contingency tables $D = (d_{ij})$ with the additional constraint that $d_{ij} = 0$ if $w_{ij} = 0$. Assuming that $\Sigma(R, C; W)$ is not empty, we consider this set as a finite probability space with the uniform measure. We call matrix W the *pattern*. We are interested in what a random table $D \in \Sigma(R, C; W)$ looks like. We define the maximum entropy matrix as before.

2.1. The maximum entropy matrix

Suppose that the set $\Sigma(R, C; W)$ is non-empty. Let us consider the polytope $\mathcal{P}(R, C; W)$ of $m \times n$ matrices $X = (x_{ij})$ such that

$$\sum_{j=1}^{n} x_{ij} = r_i \quad \text{for } i = 1, \ldots, m, \qquad \sum_{i=1}^{m} x_{ij} = c_j \quad \text{for } j = 1, \ldots, n,$$

$$0 \leqslant x_{ij} \leqslant 1 \quad \text{for all } i, j \quad \text{and} \quad x_{ij} = 0 \quad \text{whenever } w_{ij} = 0.$$

Thus $\mathcal{P}(R, C; W)$ is a face of polytope $\mathcal{P}(R, C)$ of Section 1.1.

Let $H(X)$ be the entropy function of Section 1.1. Since $H(X)$ is strictly concave, it attains a unique maximum $Z = Z(R, C; W)$ on polytope $\mathcal{P}(R, C; W)$, which we call the *maximum entropy matrix* with margins (R, C) and pattern W.

Lemma 6. *Suppose that the polytope $\mathcal{P}(R, C; W)$ contains a matrix $Y = (y_{ij})$ such that $0 < y_{ij} < 1$ whenever $w_{ij} = 1$, in which case we say that $\mathcal{P}(R, C; W)$ has a non-empty interior. Then the infimum $\alpha(R, C; W)$ in Theorem 5 is attained at a certain point $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$. The maximum entropy matrix $Z = (z_{ij})$ satisfies*

$$z_{ij} = \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} \quad \textit{for all } i, j \textit{ such that } w_{ij} = 1. \tag{3}$$

Moreover,

$$\alpha(R, C; W) = e^{H(Z)}. \tag{4}$$

Conversely, if the infimum $\alpha(R, C; W)$ is attained at a point $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$, then for the maximum entropy matrix $Z = (z_{ij})$ Eqs. (3) *and* (4) *hold.*

For $\mathcal{P}(R, C; W)$ to have a non-empty interior is equivalent to the requirement that for every pair k, l such that $w_{kl} = 1$ there is a matrix $D^0 \in \Sigma(R, C; W)$, $D^0 = (d^0_{ij})$, such that $d^0_{kl} = 0$ and there is a matrix $D^1 \in \Sigma(R, C; W)$, $D^1 = (d^1_{ij})$, such that $d^1_{kl} = 1$. In other words, we require the set $\Sigma(R, C; W)$ to be reasonably large.

We prove an analogue of Theorem 3. We consider subsets

$$S \subset \bigl\{(i, j)\colon\ w_{ij} = 1\bigr\}.$$

As before, we denote by $\sigma_S(A)$ the sum of the entries of a matrix A indexed by the subset S.

Theorem 7. *Let us fix numbers $\kappa > 0$ and $0 < \delta < 1$. Then there exists a number $q = q(\kappa, \delta)$ such that the following holds.*

Let (R, C) be margins such that $n \geqslant m > q$ and the polytope $\mathcal{P}(R, C; W)$ has a non-empty interior, and let $Z \in \mathcal{P}(R, C; W)$ be the maximum entropy matrix. Let $S \subset \{(i, j)\colon w_{ij} = 1\}$ be a subset such that $\sigma_S(Z) \geqslant \delta mn$ and let

$$\epsilon = \delta \frac{\ln n}{\sqrt{m}}.$$

If $\epsilon \leqslant 1$ then

$$\mathbf{Pr}\bigl\{D \in \Sigma(R, C; W)\colon\ (1 - \epsilon)\sigma_S(Z) \leqslant \sigma_S(D) \leqslant (1 + \epsilon)\sigma_S(Z)\bigr\} \geqslant 1 - 2n^{-\kappa n}.$$

The statement of the theorem is, of course, vacuous unless pattern W contains $\Omega(mn)$ ones.

There is an analogue of Theorem 4.

Theorem 8. *Suppose that the polytope $\mathcal{P}(R, C; W)$ has a non-empty interior and let $Z \in \mathcal{P}(R, C; W)$ be the maximum entropy matrix. Let $X = (x_{ij})$ be the random $m \times n$ matrix of independent Bernoulli random variables such that*

$$\mathbf{E}X = Z,$$

that is, $\mathbf{Pr}\{x_{ij} = 1\} = z_{ij}$ and $\mathbf{Pr}\{x_{ij} = 0\} = 1 - z_{ij}$ independently for all i, j. Then the probability mass function of X is constant on the set $\Sigma(R, C; W)$ and, moreover,

$$\mathbf{Pr}\{X = D\} = e^{-H(Z)} \quad \text{for all } D \in \Sigma(R, C; W).$$

3. Comparisons with the literature

There is a vast literature on 0–1 matrices with prescribed row and column sums and with or without zeros in prescribed positions, see, for example, [25, Chapter 16], [24,5,12], recent [8,13, 7,14] and references therein. A simple and efficient criterion for the existence of a 0–1 matrix with prescribed row and column sums is given by the classical Gale–Ryser Theorem; in the case of enforced zeros, the question reduces to the existence of a network flow, see, for example, [25, Chapter 16]. Estimating the number of such matrices also attracted a lot of attention. Precise asymptotic formulas for the number of matrices were obtained in sparse cases for which $r_i \ll n$ and $c_j \ll m$ [24,5,13], the regular case of all row sums r_i equal and all column sums c_j equal [7] and cases close to regular [7,14]. Formulas of Theorems 1 and 5 are not as precise as those of [5,7,13,14,24] but they are applicable to a wide class of margins (R, C) and they uncover some interesting features of the numbers $|\Sigma(R, C)|$ and $|\Sigma(R, C; W)|$.

The following construction provides some insight into the combinatorial interpretation of the number $\alpha(R, C)$ from Theorem 1.

3.1. *Cloning the margins*

Let us fix some margins R, C for which the set $\Sigma(R, C)$ is not empty, and, moreover, the polytope $\mathcal{P}(R, C)$ contains an interior point, so the conditions of Lemma 2 are satisfied. Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$. For a positive integer k, let us define the km-vector

$$R_k = \bigl(\underbrace{kr_1, \ldots, kr_1}_{k \text{ times}}, \ldots, \underbrace{kr_m, \ldots, kr_m}_{k \text{ times}}\bigr)$$

and the kn-vector

$$C_k = \bigl(\underbrace{kc_1, \ldots, kc_1}_{k \text{ times}}, \ldots, \underbrace{kc_n, \ldots, kc_n}_{k \text{ times}}\bigr).$$

In other words, we obtain margins (R_k, C_k) if we choose a matrix $Y \in \mathcal{P}(R, C)$ and then create a new block matrix Y_k by arranging k^2 copies of Y into a $km \times kn$ matrix. Then R_k is the vector of row sums of Y_k and C_k is the vector of column sums of Y_k. Clearly, the conditions of Lemma 2 are satisfied for (R_k, C_k).

Theorem 1 then implies that

$$\lim_{k\to+\infty}\left|\Sigma(R_k,C_k)\right|^{1/k^2}=\alpha(R,C). \tag{5}$$

Indeed, the infimum $\alpha(R,C)$ is attained at a certain point

$$\mathbf{x}^*=(\xi_1,\dots,\xi_m) \quad \text{and} \quad \mathbf{y}^*=(\eta_1,\dots,\eta_n).$$

It is not hard to see that the infimum $\alpha(R_k,C_k)$ is attained at

$$\mathbf{x}_k^*=\bigl(\underbrace{\xi_1,\dots,\xi_1}_{k\ \text{times}},\dots,\underbrace{\xi_m,\dots,\xi_m}_{k\ \text{times}}\bigr) \quad \text{and} \quad \mathbf{y}_k^*=\bigl(\underbrace{\eta_1,\dots,\eta_1}_{k\ \text{times}},\dots,\underbrace{\eta_n,\dots,\eta_n}_{k\ \text{times}}\bigr).$$

3.2. *Asymptotic repulsion in the space of matrices*

A natural candidate for an approximation of $|\Sigma(R,C)|$ is the "independence estimate"

$$I(R,C)=\binom{mn}{N}^{-1}\prod_{i=1}^{m}\binom{n}{r_i}\prod_{j=1}^{n}\binom{m}{c_j}, \tag{6}$$

see [12,13,7].

The intuitive meaning of (6) is as follows. Let us consider the set of all $m\times n$ matrices with 0–1 entries and with the total sum of entries equal to N as a finite probability space with the uniform measure. Let us consider the two events in this space: the event $\mathcal{R}$ consisting of the matrices with row sums R and the event $\mathcal{C}$ consisting of the matrices with column sums C. One can see that

$$\mathbf{Pr}(\mathcal{R})=\binom{mn}{N}^{-1}\prod_{i=1}^{m}\binom{n}{r_i} \quad \text{and} \quad \mathbf{Pr}(\mathcal{C})=\binom{mn}{N}^{-1}\prod_{j=1}^{n}\binom{m}{c_j}$$

and that

$$\left|\Sigma(R,C)\right|=\binom{mn}{N}\mathbf{Pr}(\mathcal{R}\cap\mathcal{C}).$$

Thus the value of (6) equals $|\Sigma(R,C)|$ if the events $\mathcal{R}$ and $\mathcal{C}$ are independent. It turns out that (6) indeed approximates $|\Sigma(R,C)|$ reasonably well in the sparse and near-uniform cases, see [13] and [7].

However, for generic R and C, the independence estimate $I(R,C)$ overestimates $|\Sigma(R,C)|$ by a $2^{\Omega(mn)}$ factor. To see why, let us fix some margins $R=(r_1,\dots,r_m)$ and $C=(c_1,\dots,c_n)$ such that not all row sums r_i are equal and not all column sums c_j are equal and the conditions of Lemma 2 are satisfied. Let us consider the cloned margins R_k and C_k as in Section 3.1.

Applying Stirling's formula, we get

$$\lim_{k \to +\infty} I(R_k, C_k)^{1/k^2} = \exp\left\{-mnH\left(\frac{N}{mn}\right) + n\sum_{i=1}^{m} H\left(\frac{r_i}{n}\right) + m\sum_{j=1}^{n} H\left(\frac{c_j}{m}\right)\right\}, \tag{7}$$

where H is the entropy function, see Section 1.1. To compare (7) and (5), we use Lemma 2 and the multivariate entropy function

$$\mathbf{H}(p_1, \ldots, p_k) = \sum_{i=1}^{k} p_k \ln \frac{1}{p_k},$$

where $p_1, \ldots, p_k$ are non-negative numbers such that $p_1 + \cdots + p_k = 1$. Thus $H(x) = \mathbf{H}(x, 1-x)$ for $0 \leqslant x \leqslant 1$ and we rewrite (7) as

$$\begin{aligned}\lim_{k \to +\infty} \frac{1}{k^2} \ln I(R_k, C_k) = {} & N\mathbf{H}\left(\frac{r_1}{N}, \ldots, \frac{r_m}{N}\right) + (mn-N)\mathbf{H}\left(\frac{n-r_1}{mn-N}, \ldots, \frac{n-r_m}{mn-N}\right) \\ & + N\mathbf{H}\left(\frac{c_1}{N}, \ldots, \frac{c_n}{N}\right) + (mn-N)\mathbf{H}\left(\frac{m-c_1}{mn-N}, \ldots, \frac{m-c_n}{mn-N}\right) \\ & - N\ln N - (mn-N)\ln(mn-N).\end{aligned}$$

On the other hand, applying Lemma 2, we can rewrite (5) as

$$\begin{aligned}\lim_{k \to +\infty} \frac{1}{k^2} \ln\left|\Sigma(R_k, C_k)\right| = {} & N\mathbf{H}\left(\frac{z_{ij}}{N}\right) + (mn-N)\mathbf{H}\left(\frac{1-z_{ij}}{mn-N}\right) - N\ln N \\ & - (mn-N)\ln(mn-N),\end{aligned}$$

where $Z = (z_{ij})$ is the maximum entropy matrix for margins (R, C).

We now use some classical entropy inequalities, see, for example, [19]. Namely, by the inequality relating the entropies of two partitions of a probability space and the entropy of their intersection, we have

$$\mathbf{H}\left(\frac{z_{ij}}{N}\right) \leqslant \mathbf{H}\left(\frac{r_1}{N}, \ldots, \frac{r_m}{N}\right) + \mathbf{H}\left(\frac{c_1}{N}, \ldots, \frac{c_n}{N}\right)$$

with the equality if and only if

$$z_{ij} = \frac{r_i c_j}{N} \quad \text{for all } i, j \tag{8}$$

and

$$\mathbf{H}\left(\frac{1-z_{ij}}{mn-N}\right) \leqslant \mathbf{H}\left(\frac{n-r_1}{mn-N}, \ldots, \frac{n-r_m}{mn-N}\right) + \mathbf{H}\left(\frac{m-c_1}{mn-N}, \ldots, \frac{m-c_n}{mn-N}\right)$$

with the equality if and only if

$$1 - z_{ij} = \frac{(n - r_i)(m - c_j)}{mn - N} \quad \text{for all } i, j. \tag{9}$$

However, if we have both (8) and (9), we must have $(r_i m - N)(c_j n - N) = 0$, so unless all row sums r_i are equal or all column sums c_j are equal, we have

$$\lim_{k \to +\infty} \left| \Sigma(R_k, C_k) \right|^{1/k^2} < \lim_{k \to +\infty} I(R_k, C_k)^{1/k^2}.$$

Therefore, as k grows, the independence estimate (6) overestimates the number of 0–1 matrices with row sums R_k and column sums C_k by a factor of $2^{\Omega(k^2)}$. In probabilistic terms, as k grows, the event $\mathcal{R}_k$ consisting of the 0–1 matrices with row sums R_k and the event $\mathcal{C}_k$ consisting of the 0–1 matrices with column sums C_k repel each other (the events are negatively correlated), instead of being asymptotically independent.

The procedure of cloning described in Section 3.1 produces margins of increasing size with the following features: the density remains separated from 0 and 1, and if the margins were non-uniform initially, they stay away from uniform. One can show that for more general sequences of margins that share these two features, we have the asymptotic repulsion of the event consisting of the 0–1 matrices with prescribed row sums and the event consisting of the 0–1 matrices with prescribed column sums. This is in contrast to the case of contingency tables (non-negative integer matrices with prescribed row and column sums), where we have the asymptotic attraction of the events [3].

3.3. *Randomized counting and sampling*

Jerrum, Sinclair, and Vigoda [18] showed how to apply their algorithm for computing the permanent of a non-negative matrix to construct a fully polynomial randomized approximation scheme (FPRAS) to compute $|\Sigma(R, C)|$ and, more generally, $|\Sigma(R, C; W)|$, where W is a 0–1 pattern, see also [6]. Furthermore, they obtained a polynomial time algorithm for sampling a random $D \in \Sigma(R, C)$ and $D \in \Sigma(R, C; W)$ from a "nearly uniform" distribution. This problem arises naturally in statistics, see, for example, [9]. The estimates of Theorems 1 and 5 are not nearly as precise as those of [18], but they are deterministic, easily computable, and amenable to analysis. Similarly, we do not provide a sampling algorithm but show instead in Theorems 3 and 7 what a random matrix is likely to look like.

3.4. *An open question*

Theorem 4 allows us to interpret Theorem 3 as a law of large numbers for binary contingency tables: with respect to sums $\sigma_S(D)$ for sufficiently "heavy" sets S of indices, a random binary contingency table $D \in \Sigma(R, C)$ behaves approximately as the matrix of independent Bernoulli random variables whose expectation is the maximum entropy matrix $Z = (z_{ij})$. Similar concentration results can be obtained for other well-behaved functions on binary contingency tables. One can ask whether the distribution of a *particular entry* d_{ij} of a random table $D \in \Sigma(R, C)$ converges in distribution to the Bernoulli distribution with expectation z_{ij} as the dimensions m and n of the table grow in some regular way, for example, when the margins are cloned as in Section 3.1.

Our approach, based on estimating combinatorial quantities via solutions to optimization problems, reminds one of that of Gurvits [16]. The appearance of entropy in combinatorial counting problems reminds one of recent papers of Cuckler and Kahn [10,11], although methods and results seem to be quite different.

In the rest of the paper, we prove the results stated in Sections 1 and 2.

4. Preliminaries: permanents and scaling

Let $A = (a_{ij})$ be an $n \times n$ matrix. The *permanent* of A is defined by the expression

$$\operatorname{per} A = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)},$$

where S_n is the symmetric group of all permutations of the set $\{1, \dots, n\}$. The relevance of permanents to us is that both values of $|\Sigma(R, C)|$ and $|\Sigma(R, C; W)|$ can be expressed as permanents of $mn \times mn$ matrices. This result is not new, for $|\Sigma(R, C)|$ it was observed, for example, in [17]. For $|\Sigma(R, C; W)|$, where W is a 0–1 pattern, a construction is presented in [18]. We give a general construction for $|\Sigma(R, C; W)|$, where W is an arbitrary matrix, which is slightly different from that of [18].

Lemma 9. *Let us choose margins $R = (r_1, \dots, r_m)$, $C = (c_1, \dots, c_n)$ and an $m \times n$ matrix $W = (w_{ij})$ of weights. Let us construct an $mn \times mn$ matrix $A = A(R, C; W)$ as follows.*

The rows of A are split into disjoint m blocks having $n - r_1, \dots, n - r_m$ rows respectively (blocks of type I) and n blocks having $c_1, \dots, c_n$ rows respectively (blocks of type II).

The columns of A are split into m disjoint blocks of n columns in each.

For $i = 1, \dots, m$ the entry of A that lies in a row from the i-th block of rows of type I and in a column from the i-th block of columns is equal to 1.

For $i = 1, \dots, m$ and $j = 1, \dots, n$ the entry of A that lies in a row from the j-th block of rows of type II and the j-th column from the i-th block of columns is equal to w_{ij}.

All other entries of A are 0*s.*

Then

$$\left|\Sigma(R, C; W)\right| = \left(\prod_{i=1}^{m} \frac{1}{(n - r_i)!}\right)\left(\prod_{j=1}^{n} \frac{1}{c_j!}\right) \operatorname{per} A.$$

Proof. First, we express $|\Sigma(R, C; W)|$ as a coefficient in a certain polynomial. Let $x_1, \dots, x_n$ be formal variables and let

$$e_r(x_1, \dots, x_n) = \sum_{1 \leqslant i_1 < \dots < i_r \leqslant n} x_{i_1} \cdots x_{i_r}$$

be the elementary symmetric polynomial of degree r. Thus $e_r(x_1, \dots, x_n)$ is

$$\text{the coefficient of } t^{n-r} \text{ in the product } \prod_{j=1}^{n} (t + x_j).$$

We observe that $|\Sigma(R, C; W)|$ is

$$\text{the coefficient of } \prod_{j=1}^{n} x_j^{c_j} \text{ in the product } \prod_{i=1}^{m} e_{r_i}(w_{i1}x_1, \dots, w_{in}x_n).$$

Summarizing, we conclude that $|\Sigma(R, C; W)|$ is

$$\text{the coefficient of } \prod_{i=1}^{m} t_i^{n-r_i} \prod_{j=1}^{n} x_j^{c_j} \text{ in the product } \prod_{i=1}^{m}\prod_{j=1}^{n}(t_i + w_{ij}x_j).$$

To express the last coefficient as the permanent of a matrix, we use a convenient scalar product in the space of polynomials, see, for example, [1] and [2]. Namely, for monomials

$$\mathbf{x}^a = x_1^{\alpha_1} \cdots x_n^{\alpha_n} \quad \text{where } a = (\alpha_1, \dots, \alpha_n) \text{ and } \mathbf{x} = (x_1, \dots, x_n)$$

we define

$$\langle \mathbf{x}^a, \mathbf{x}^b \rangle = \begin{cases} \alpha_1! \cdots \alpha_n! & \text{if } a = b = (\alpha_1, \dots, \alpha_n), \\ 0 & \text{if } a \neq b \end{cases}$$

and then extend the scalar product $\langle \cdot, \cdot \rangle$ by bilinearity. Equivalently, the scalar product can be defined as follows: let us identify $\mathbb{R}^n \oplus \mathbb{R}^n = \mathbb{C}^n$ via $x + iy = z$ and let ν_n be the Gaussian measure on $\mathbb{C}^n$ with the density

$$\pi^{-n} e^{-\|z\|^2} \quad \text{where } \|z\|^2 = \|x\|^2 + \|y\|^2 \text{ for } z = x + iy.$$

Then, for polynomials f and g we have

$$\langle f, g \rangle = \int_{\mathbb{C}^n} f(z)\overline{g(z)}\, d\nu_n,$$

where $\bar{g}$ is the complex conjugate of g, see, for example, [2, Section 4].

The convenient property of the scalar product is that if

$$p(\mathbf{x}) = \prod_{l=1}^{m} \sum_{k=1}^{n} b_{lk} x_k \quad \text{and} \quad q(\mathbf{x}) = \prod_{l=1}^{m} \sum_{k=1}^{n} c_{lk} x_k$$

are products of linear forms, then

$$\langle p, q \rangle = \operatorname{per} D,$$

where $D = (d_{ij})$ is the $m \times m$ matrix defined by

$$d_{ij} = \sum_{k=1}^{n} b_{ik} c_{jk} \quad \text{for all } i, j,$$

see [2, Lemma 4.5] or, for a more general identity, [15, Theorem 3.8]. Thus we may write

$$\begin{aligned}\left|\Sigma(R,C;W)\right| &= \left(\prod_{i=1}^{m}\frac{1}{(n-r_i)!}\right)\left(\prod_{j=1}^{n}\frac{1}{c_j!}\right)\left\langle\prod_{i=1}^{m}t_i^{n-r_i}\prod_{j=1}^{n}x_j^{c_j},\prod_{i=1}^{m}\prod_{j=1}^{n}(t_i+w_{ij}x_j)\right\rangle\\ &= \left(\prod_{i=1}^{m}\frac{1}{(n-r_i)!}\right)\left(\prod_{j=1}^{n}\frac{1}{c_j!}\right)\operatorname{per}A. \qquad \square\end{aligned}$$

4.1. Matrix scaling and the van der Waerden bound

Let $B=(b_{ij})$ be an $n\times n$ matrix. Matrix B is called *doubly stochastic* if

$$\sum_{j=1}^{n}b_{ij}=1 \quad \text{for } i=1,\dots,m, \qquad \sum_{i=1}^{n}b_{ij}=1 \quad \text{for } j=1,\dots,n$$
$$\text{and} \quad b_{ij}\geqslant 0 \quad \text{for all } i, j.$$

The classical bound conjectured by van der Waerden and proved by Falikman and Egorychev, see [25, Chapter 12] and also [16] for exciting new developments, states that

$$\operatorname{per}B\geqslant\frac{n!}{n^n}$$

if B is a doubly stochastic matrix.

Linial, Samorodnitsky, and Wigderson [20] introduced the following very useful *scaling* method of approximating permanents of non-negative matrices. Given a non-negative $n\times n$ matrix $A=(a_{ij})$ one finds non-negative numbers $\lambda_1,\dots,\lambda_n$ and $\mu_1,\dots,\mu_n$ and a doubly stochastic matrix $B=(b_{ij})$ such that

$$a_{ij}=\lambda_i\mu_j b_{ij} \quad \text{for all } i, j.$$

Then

$$\operatorname{per}A=\left(\prod_{i=1}^{n}\lambda_i\right)\left(\prod_{j=1}^{n}\mu_j\right)\operatorname{per}B$$

and an estimate of $\operatorname{per}B$ (such as the van der Waerden estimate) implies an estimate of $\operatorname{per}A$. If A is strictly positive, such doubly stochastic matrix B and scaling factors λ_i, μ_j always exist. In our situation, matrix A constructed in Lemma 9 is only non-negative. We will not always be able to scale it to a doubly stochastic matrix B exactly, but we will scale it *approximately*.

We restate a weaker form of [20, Proposition 5.1] regarding almost doubly stochastic matrices.

Lemma 10. *For any n there exists an $\epsilon_0 = \epsilon_0(n) > 0$ and a function $\phi(\epsilon)$, $0 < \epsilon < \epsilon_0$, such that*

$$\lim_{\epsilon \to 0+} \phi(\epsilon) = 1$$

and for any $n \times n$ non-negative matrix $B = (b_{ij})$ such that

$$\sum_{i=1}^{n} b_{ij} = 1 \quad \text{for } j = 1, \ldots, n$$

and

$$1 - \epsilon \leqslant \sum_{j=1}^{n} b_{ij} \leqslant 1 + \epsilon \quad \text{for } i = 1, \ldots, n$$

for some $0 \leqslant \epsilon < \epsilon_0$, we have

$$\operatorname{per} B \geqslant \frac{n!}{n^n} \phi(\epsilon).$$

From [20], one can choose $\epsilon_0 = 1/n$ and $\phi(\epsilon) = (1 - \epsilon n)^n$.

5. Proofs of Theorems 1 and 5

We prove Theorem 5 only since Theorem 1 is a particular case of Theorem 5. We start with a straightforward observation.

Lemma 11. *We have*

$$\prod_{ij} (1 + w_{ij} x_i y_j) = \sum_{(R,C)} \left| \Sigma(R, C; W) \right| \mathbf{x}^R \mathbf{y}^C,$$
$$\text{where } \mathbf{x}^R = x_1^{r_1} \cdots x_m^{r_m}, \ \mathbf{y}^C = y_1^{c_1} \cdots y_n^{c_n},$$

and the sum is taken over all margins R, C.

Next, we need a technical lemma.

Lemma 12. *Let $W = (w_{ij})$ be an $m \times n$ non-negative matrix such that*

$$\alpha(R, C; W) > 0.$$

Then, for any $\epsilon > 0$ there exist points $\mathbf{x} = \mathbf{x}(\epsilon)$ and $\mathbf{y} = \mathbf{y}(\epsilon)$, $\mathbf{x} = (x_1, \ldots, x_m)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, such that

$$\left| -r_i + \sum_{j=1}^{n} \frac{w_{ij} x_i y_j}{1 + w_{ij} x_i y_j} \right| < \epsilon \quad \textit{for } i = 1, \ldots, m,$$

$$\left| -c_j + \sum_{i=1}^{m} \frac{w_{ij} x_i y_j}{1 + w_{ij} x_i y_j} \right| < \epsilon \quad \textit{for } j = 1, \ldots, n \quad \textit{and} \quad x_i, y_j > 0 \quad \textit{for all } i, j.$$

Proof. Let us consider the function

$$G(\mathbf{s}, \mathbf{t}; W) = -\sum_{i=1}^{m} r_i s_i - \sum_{j=1}^{n} c_j t_j + \sum_{ij} \ln\bigl(1 + w_{ij} e^{s_i + t_j}\bigr)$$
$$\text{for } \mathbf{s} = (s_1, \ldots, s_m) \text{ and } \mathbf{t} = (t_1, \ldots, t_n).$$

Then $G(\mathbf{s}, \mathbf{t}; W)$ is convex and

$$\inf_{\substack{\mathbf{s} \in \mathbb{R}^m \\ \mathbf{t} \in \mathbb{R}^n}} G(\mathbf{s}, \mathbf{t}) = \ln \alpha(R, C; W) > -\infty.$$

Hence $G(\mathbf{s}, \mathbf{t})$ is bounded from below, it is also easy to check that the Hessian of G remains bounded on $\mathbb{R}^m \times \mathbb{R}^n$. Therefore, the gradient of $G(\mathbf{s}, \mathbf{t})$ can get arbitrarily close to 0. That is, for any $\epsilon > 0$ there are points

$$\mathbf{s}(\epsilon) = \bigl(s_1(\epsilon), \ldots, s_m(\epsilon)\bigr) \quad \text{and} \quad \mathbf{t}(\epsilon) = \bigl(t_1(\epsilon), \ldots, t_n(\epsilon)\bigr)$$

such that

$$\left| \frac{\partial}{\partial s_i} G(\mathbf{s}, \mathbf{t})|_{\mathbf{s}=\mathbf{s}(\epsilon), \mathbf{t}=\mathbf{t}(\epsilon)} \right| < \epsilon \quad \text{for } i = 1, \ldots, m \quad \text{and}$$
$$\left| \frac{\partial}{\partial t_j} G(\mathbf{s}, \mathbf{t})|_{\mathbf{s}=\mathbf{s}(\epsilon), \mathbf{t}=\mathbf{t}(\epsilon)} \right| < \epsilon \quad \text{for } j = 1, \ldots, n$$

(it suffices to choose $\mathbf{s}(\epsilon)$ and $\mathbf{t}(\epsilon)$ so that the value of $G(\mathbf{s}(\epsilon), \mathbf{t}(\epsilon))$ is sufficiently close to the infimum). In other words,

$$\left| -r_i + \sum_{j=1}^{n} \frac{w_{ij} e^{s_i(\epsilon) + t_j(\epsilon)}}{1 + w_{ij} e^{s_i(\epsilon) + t_j(\epsilon)}} \right| < \epsilon \quad \text{for } i = 1, \ldots, m$$

and

$$\left| -c_j + \sum_{i=1}^{m} \frac{w_{ij} e^{s_i(\epsilon) + t_j(\epsilon)}}{1 + w_{ij} e^{s_i(\epsilon) + t_j(\epsilon)}} \right| < \epsilon \quad \text{for } j = 1, \ldots, n.$$

We now let

$$x_i = x_i(\epsilon) = e^{s_i(\epsilon)} \quad \text{for } i = 1, \ldots, m \quad \text{and}$$
$$y_j = y_j(\epsilon) = e^{t_j(\epsilon)} \quad \text{for } j = 1, \ldots, n. \qquad \square$$

5.1. Proof of Theorem 5

The upper bound

$$\alpha(R, C; W) \geqslant \left|\Sigma(R, C; W)\right|$$

follows from Lemma 11. Let us prove the lower bound.

If $\alpha(R, C; W) = 0$ then $|\Sigma(R, C; W)| = 0$ and the lower bound follows. Hence we assume that $\alpha(R, C; W) > 0$.

Let $A = A(R, C; W)$ be the $mn \times mn$ block matrix constructed in Lemma 9. Let us consider the $mn \times mn$ block matrix $B(\epsilon)$ obtained from A as follows. For $\epsilon > 0$, let $\mathbf{x}(\epsilon) = (x_1, \dots, x_m)$ and $\mathbf{y}(\epsilon) = (y_1, \dots, y_n)$ be the point constructed in Lemma 12.

For $i = 1, \dots, m$ we multiply every row of A in the i-th block of type I by

$$\frac{1}{x_i(n - r_i)}.$$

For $j = 1, \dots, n$ we multiply every row of A in the j-th block of type II by

$$\frac{y_j}{c_j} \quad \text{for } j = 1, \dots, n.$$

For $i = 1, \dots, m$ and $j = 1, \dots, n$ we multiply the j-th column in the i-th block of columns of A by

$$\frac{x_i}{1 + w_{ij} x_i y_j}.$$

This choice of scaling factors is, basically, a lucky guess made in the hope to match the structure of the function $F(\mathbf{x}, \mathbf{y}; W)$.

Thus we have

$$\operatorname{per} A = \left(\prod_{i=1}^{m} x_i^{n-r_i} (n - r_i)^{n-r_i}\right)\left(\prod_{j=1}^{n} y_j^{-c_j} c_j^{c_j}\right)\left(\prod_{ij} x_i^{-1}(1 + w_{ij} x_i y_j)\right) \operatorname{per} B(\epsilon)$$

and hence

$$\begin{aligned} \left|\Sigma(R, C; W)\right| &= \left(\prod_{i=1}^{m} \frac{(n - r_i)^{n-r_i}}{(n - r_i)!}\right)\left(\prod_{j=1}^{n} \frac{c_j^{c_j}}{c_j^{c_j}}\right) F\big(\mathbf{x}(\epsilon), \mathbf{y}(\epsilon); W\big) \operatorname{per} B(\epsilon) \\ &\geqslant \left(\prod_{i=1}^{m} \frac{(n - r_i)^{n-r_i}}{(n - r_i)!}\right)\left(\prod_{j=1}^{n} \frac{c_j^{c_j}}{c_j^{c_j}}\right) \alpha(R, C; W) \operatorname{per} B(\epsilon). \end{aligned} \tag{10}$$

Finally, we claim that $B(\epsilon)$ is close to a doubly stochastic matrix. Indeed, for $i = 1, \dots, m$ and $j = 1, \dots, n$ the entry of $B(\epsilon)$ that lies in a row from the i-th block of rows of type I and in the j-th column from the i-th block of columns is equal to

$$\frac{1}{(n-r_i)(1+w_{ij}x_i y_j)}.$$

For $i=1,\dots,m$ and $j=1,\dots,n$ the entry of $B(\epsilon)$ that lies in a row from the j-th block of rows of type II and the j-th column from the i-th block of columns is equal to

$$\frac{w_{ij}x_i y_j}{c_j(1+w_{ij}x_i y_j)}.$$

All other entries of $B(\epsilon)$ are 0s. Let us compute the row sums of $B(\epsilon)$.

For a row in the i-th block of rows of type I the sum equals

$$a_i = \sum_{j=1}^{n} \frac{1}{(n-r_i)(1+w_{ij}x_i y_j)}.$$

Since

$$\sum_{j=1}^{n} \frac{1}{1+w_{ij}x_i y_j} = \sum_{j=1}^{n} \frac{1+w_{ij}x_i y_j}{1+w_{ij}x_i y_j} - \sum_{j=1}^{n} \frac{w_{ij}x_i y_j}{1+w_{ij}x_i y_j},$$

by the inequalities of Lemma 12, we have

$$|a_i - 1| < \frac{\epsilon}{n-r_i} \leqslant \epsilon \quad \text{for } i=1,\dots,m.$$

For a row in the j-th block of rows of type II the sum equals

$$b_j = \sum_{i=1}^{m} \frac{w_{ij}x_i y_j}{c_j(1+w_{ij}x_i y_j)}.$$

By the inequalities of Lemma 12, we have

$$|b_j - 1| < \frac{\epsilon}{c_j} \leqslant \epsilon \quad \text{for } j=1,\dots,n.$$

Let us compute the column sums of $B(\epsilon)$.

For the j-th column from the i-th block of columns the sum equals

$$(n-r_i)\frac{1}{(n-r_i)(1+w_{ij}x_i y_j)} + c_j \frac{w_{ij}x_i y_j}{c_j(1+w_{ij}x_i y_j)} = 1.$$

Clearly, $B(\epsilon)$ is non-negative and hence by Lemma 10, we have

$$\operatorname{per} B(\epsilon) \geqslant \frac{(mn)!}{(mn)^{mn}}\phi(\epsilon) \quad \text{where } \lim_{\epsilon\to 0+}\phi(\epsilon)=1.$$

The proof now follows by (10) as $\epsilon \to 0+$. $\square$

6. Proofs of Lemmas 2 and 6

We prove Lemma 6 only since Lemma 2 is a particular case of Lemma 6.

Proof of Lemma 6. Since $H'(x) = \ln(1-x) - \ln x$, the value of the derivative at $x = 0$ is $+\infty$ (we consider the right derivative there), the value of the derivative at $x = 1$ is $-\infty$ (we consider the left derivative there) and the value of the derivative is finite for any $0 < x < 1$. Suppose that for the maximum entropy matrix Z we have $z_{ij} \in \{0, 1\}$ for some i, j such that $w_{ij} = 1$. If $Y \in \mathcal{P}(R, C; W)$, $Y = (y_{ij})$, is a matrix such that $0 < y_{ij} < 1$ whenever $w_{ij} = 1$ then

$$H\big(\epsilon Y + (1-\epsilon)Z\big) > H(Z) \quad \text{for a sufficiently small } \epsilon > 0,$$

which contradicts to the choice of Z. Hence

$$0 < z_{ij} < 1 \quad \text{whenever } w_{ij} = 1.$$

Therefore, the gradient of $H(X)$ at $X = Z$ is orthogonal to the affine subspace of matrices $X = (x_{ij})$ having row sums R, column sums C, and such that $x_{ij} = 0$ whenever $w_{ij} = 0$. Hence

$$\ln \frac{1 - z_{ij}}{z_{ij}} = \lambda_i + \mu_j \quad \text{for all } i, j \text{ such that } w_{ij} = 1 \tag{11}$$

and some $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$. Hence

$$z_{ij} = \frac{e^{-\lambda_i} e^{-\mu_j}}{1 + e^{-\lambda_i} e^{-\mu_j}} \quad \text{whenever } w_{ij} = 1.$$

Therefore

$$\sum_{j:\ w_{ij}=1} \frac{e^{-\lambda_i} e^{-\mu_j}}{1 + e^{-\lambda_i} e^{-\mu_j}} = r_i \quad \text{for } i = 1, \ldots, m,$$

$$\sum_{i:\ w_{ij}=1} \frac{e^{-\lambda_i} e^{-\mu_j}}{1 + e^{-\lambda_i} e^{-\mu_j}} = c_j \quad \text{for } j = 1, \ldots, n.$$

Therefore,

$$\mathbf{s}^* = (-\lambda_1, \ldots, -\lambda_m) \quad \text{and} \quad \mathbf{t}^* = (-\mu_1, \ldots, -\mu_n)$$

is a critical point of

$$G(\mathbf{s}, \mathbf{t}; W) = -\sum_{i=1}^m r_i s_i - \sum_{j=1}^n c_j t_j + \sum_{(i,j):\ w_{ij}=1} \ln\big(1 + e^{s_i + t_j}\big).$$

Since G is convex, $(\mathbf{s}^*, \mathbf{t}^*)$ is also a minimum point. Therefore, the point $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$ where

$$\xi_i = e^{-\lambda_i} \quad \text{for } i = 1, \ldots, m \quad \text{and} \quad \eta_j = e^{-\mu_j} \quad \text{for } j = 1, \ldots, n$$

is a minimum point of

$$F(\mathbf{x}, \mathbf{y}; W) = \left(\prod_{i=1}^{m} x_i^{-r_i}\right)\left(\prod_{j=1}^{n} y_j^{-c_j}\right) \prod_{(i,j):\ w_{ij}=1} (1 + x_i y_j)$$

and satisfies

$$\begin{aligned} \sum_{j:\ w_{ij}=1} \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} &= r_i \quad \text{for } i = 1, \ldots, m, \\ \sum_{i:\ w_{ij}=1} \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} &= c_j \quad \text{for } j = 1, \ldots, n. \end{aligned} \tag{12}$$

Conversely, if $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$ is a point where the minimum of $F(\mathbf{x}, \mathbf{y}; W)$ is attained, then, setting the gradient of $\ln F$ to 0, we obtain Eqs. (12). Letting

$$z_{ij} = \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} \quad \text{when } w_{ij} = 1$$

and $z_{ij} = 0$ when $w_{ij} = 0$, we obtain a matrix $Z \in \mathcal{P}(R, C; W)$. Moreover, the gradient of $H(X)$ at $X = Z$ satisfies (11) with $\lambda_i = -\ln \xi_i$ and $\mu_j = -\ln \eta_j$, so Z is the maximum entropy matrix.

We now check:

$$\begin{aligned} H(Z) &= -\sum_{(i,j):\ w_{ij}=1} z_{ij} \ln z_{ij} - \sum_{(i,j):\ w_{ij}=1} (1 - z_{ij}) \ln(1 - z_{ij}) \\ &= -\sum_{(i,j):\ w_{ij}=1} \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} \ln \frac{\xi_i \eta_j}{1 + \xi_i \eta_j} - \sum_{(i,j):\ w_{ij}=1} \frac{1}{1 + \xi_i \eta_j} \ln \frac{1}{1 + \xi_i \eta_j} \\ &= -\sum_{i=1}^{m} \ln \xi_i \left(\sum_{j:\ w_{ij}=1} \frac{\xi_i \eta_j}{1 + \xi_i \eta_j}\right) - \sum_{j=1}^{n} \ln \eta_j \left(\sum_{i:\ w_{ij}=1} \frac{\xi_i \eta_j}{1 + \xi_i \eta_j}\right) \\ &\quad + \sum_{(i,j):\ w_{ij}=1} \ln(1 + \xi_i \eta_j) \\ &= -\sum_{i=1}^{m} r_i \ln \xi_i - \sum_{j=1}^{n} c_j \ln \eta_j + \sum_{(i,j):\ w_{ij}=1} \ln(1 + \xi_i \eta_j) \end{aligned}$$

by (12). Hence

$$H(Z) = \ln F\big(\mathbf{x}^*, \mathbf{y}^*; W\big)$$

and the proof follows. □

7. Proofs of Theorems 4 and 8

We prove Theorem 8 only, since Theorem 4 is a particular case of Theorem 8.

From formula (11), we have

$$\frac{1-z_{ij}}{z_{ij}} = e^{\lambda_i+\mu_j} \quad \text{for all } i, j \text{ such that } w_{ij}=1$$

and some $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$. Then, for all i, j such that $w_{ij}=1$ and any $d_{ij} \in \{0,1\}$, we have

$$\begin{aligned}\mathbf{Pr}\{x_{ij}=d_{ij}\} &= z_{ij}^{d_{ij}}(1-z_{ij})^{1-d_{ij}} = (1-z_{ij})\left(\frac{1-z_{ij}}{z_{ij}}\right)^{-d_{ij}} \\ &= (1-z_{ij})e^{-(\lambda_i+\mu_j)d_{ij}}.\end{aligned}$$

Consequently, for any $D \in \Sigma(R, C; W)$, $D = (d_{ij})$, we have

$$\begin{aligned}\mathbf{Pr}\{X=D\} &= \prod_{i,j:\ w_{ij}=1} (1-z_{ij})e^{-(\lambda_i+\mu_j)d_{ij}} \\ &= \left(\prod_{i,j:\ w_{ij}=1}(1-z_{ij})\right)\left(\prod_{i=1}^{m} e^{-\lambda_i r_i}\right)\left(\prod_{j=1}^{n} e^{-\mu_j c_j}\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}e^{-H(Z)} &= \prod_{i,j:\ w_{ij}=1} z_{ij}^{z_{ij}}(1-z_{ij})^{1-z_{ij}} \\ &= \left(\prod_{i,j:\ w_{ij}=1}(1-z_{ij})\right)\left(\prod_{i,j:\ w_{ij}=1}\left(\frac{1-z_{ij}}{z_{ij}}\right)^{-z_{ij}}\right) \\ &= \left(\prod_{i,j:\ w_{ij}=1}(1-z_{ij})\right)\left(\prod_{i=1}^{m} e^{-\lambda_i r_i}\right)\left(\prod_{j=1}^{n} e^{-\mu_j c_j}\right),\end{aligned}$$

which completes the proof. □

8. Proofs of Theorems 3 and 7

We prove Theorem 7 only since Theorem 3 is a particular case of Theorem 7.

We will use standard large deviation inequalities for bounded random variables, see, for example, [22, Corollary 5.3].

Lemma 13. *Let $Y_1, \ldots, Y_k$ be independent random variables such that $0 \leqslant Y_i \leqslant 1$ for $i = 1, \ldots, k$. Let $Y = Y_1 + \cdots + Y_k$ and let $a = \mathbf{E}Y$. Then, for $0 \leqslant \epsilon \leqslant 1$ we have*

$$\mathbf{Pr}\{Y \geqslant (1+\epsilon)a\} \leqslant \exp\left\{-\frac{1}{3}\epsilon^2 a\right\} \quad \textit{and} \quad \mathbf{Pr}\{Y \leqslant (1-\epsilon)a\} \leqslant \exp\left\{-\frac{1}{2}\epsilon^2 a\right\}.$$

8.1. Proof of Theorem 7

Let $X = (x_{ij})$ be the $m \times n$ matrix of independent Bernoulli random variables such that $\mathbf{E}X = Z$, as in Theorem 8. By Theorem 8, the distribution of X conditioned on $\Sigma(R, C; W)$ is uniform and hence

$$\begin{aligned}&\mathbf{Pr}\{D \in \Sigma(R, C; W)\colon \sigma_S(D) \leqslant (1-\epsilon)\sigma_S(Z)\}\\ &\quad = \frac{\mathbf{Pr}\{X\colon \sigma_S(X) \leqslant (1-\epsilon)\sigma_S(Z) \text{ and } X \in \Sigma(R, C; W)\}}{\mathbf{Pr}\{X\colon X \in \Sigma(R, C; W)\}}.\end{aligned}$$

Similarly,

$$\begin{aligned}&\mathbf{Pr}\{D \in \Sigma(R, C; W)\colon \sigma_S(D) \geqslant (1+\epsilon)\sigma_S(Z)\}\\ &\quad = \frac{\mathbf{Pr}\{X\colon \sigma_S(X) \geqslant (1+\epsilon)\sigma_S(Z) \text{ and } X \in \Sigma(R, C; W)\}}{\mathbf{Pr}\{X\colon X \in \Sigma(R, C; W)\}}.\end{aligned}$$

By Theorem 8, Lemma 6 and Theorem 5, we get

$$\begin{aligned}\mathbf{Pr}\{X \in \Sigma(R, C; W)\} &= e^{-H(Z)}\left|\Sigma(R, C; W)\right|\\ &\geqslant \frac{(mn)!}{(mn)^{mn}}\left(\prod_{i=1}^{m}\frac{(n-r_i)^{n-r_i}}{(n-r_i)!}\right)\left(\prod_{j=1}^{n}\frac{c_j^{c_j}}{c_j!}\right)\\ &\geqslant (mn)^{-\gamma(m+n)}\end{aligned}$$

for some absolute constant $\gamma > 0$.

Therefore,

$$\begin{aligned}&\mathbf{Pr}\{D \in \Sigma(R, C; W)\colon \sigma(D) \leqslant (1-\epsilon)\sigma_S(Z)\}\\ &\quad \leqslant (mn)^{\gamma(m+n)}\mathbf{Pr}\{X\colon \sigma_S(X) \leqslant (1-\epsilon)\sigma_S(Z)\}\end{aligned}$$

and similarly

$$\begin{aligned}&\mathbf{Pr}\{D \in \Sigma(R, C; W)\colon \sigma(D) \geqslant (1+\epsilon)\sigma_S(Z)\}\\ &\quad \leqslant (mn)^{\gamma(m+n)}\mathbf{Pr}\{X\colon \sigma_S(X) \geqslant (1+\epsilon)\sigma_S(Z)\}.\end{aligned} \tag{13}$$

By Lemma 13,

$$\mathbf{Pr}\{X\colon \sigma_S(X) \leqslant (1-\epsilon)\sigma_S(Z)\} \leqslant \exp\left\{-\frac{1}{2}\epsilon^2\sigma_S(Z)\right\}$$

and

$$\mathbf{Pr}\left\{X\colon \sigma_S(X) \geqslant (1+\epsilon)\sigma_S(Z)\right\} \leqslant \exp\left\{-\frac{1}{3}\epsilon^2\sigma_S(Z)\right\}. \tag{14}$$

Hence for

$$\epsilon = \delta\frac{\ln n}{\sqrt{m}} \quad \text{and} \quad \sigma_S(Z) \geqslant \delta mn$$

we have

$$\epsilon^2\sigma_S(Z) \geqslant \delta^3 n \ln^2 n. \tag{15}$$

Combining (13)–(15), we conclude that for any $\kappa > 0$ and all sufficiently large $n \geqslant m > q(\kappa, \delta)$ we have

$$\mathbf{Pr}\left\{D \in \Sigma(R, C; W)\colon \sigma_S(D) \leqslant (1-\epsilon)\sigma_S(Z)\right\} \leqslant n^{-\kappa n}$$

and

$$\mathbf{Pr}\left\{D \in \Sigma(R, C; W)\colon \sigma_S(D) \geqslant (1+\epsilon)\sigma_S(Z)\right\} \leqslant n^{-\kappa n}$$

as required. □

Acknowledgments

I am grateful to Alex Samorodnitsky for emphasizing the importance of matrix scaling in permanent computations during many enlightening conversations. I benefited from conversations with Catherine Greenhill about asymptotic enumeration during the 2008 Schloss Dagstuhl meeting on design and analysis of randomized and approximation algorithms. After the first version of this paper was written, John Hartigan pointed out to connections with the maximum entropy principle and suggested Theorem 4 to me (cf. [4]), which led to a substantial simplification of the original proof and some strengthening of Theorems 3 and 7.

References

[1] A. Barvinok, Two algorithmic results for the traveling salesman problem, Math. Oper. Res. 21 (1996) 65–84.
[2] A. Barvinok, Integration and optimization of multivariate polynomials by restriction onto a random subspace, Found. Comput. Math. 7 (2007) 229–244.
[3] A. Barvinok, Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes, Int. Math. Res. Not. IMRN 2009 (2009) 348–385.
[4] A. Barvinok, J. Hartigan, Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes, preprint arXiv:0903.5223, 2009.
[5] E.A. Bender, The asymptotic number of non-negative integer matrices with given row and column sums, Discrete Math. 10 (1974) 217–223.
[6] I. Bezáková, N. Bhatnagar, E. Vigoda, Sampling binary contingency tables with a greedy start, Random Structures Algorithms 30 (2007) 168–205.
[7] E.R. Canfield, C. Greenhill, B.D. McKay, Asymptotic enumeration of dense 0–1 matrices with specified line sums, J. Combin. Theory Ser. A 115 (2008) 32–66.
[8] E.R. Canfield, B.D. McKay, Asymptotic enumeration of dense 0–1 matrices with equal row sums and equal column sums, Electron. J. Combin. 12 (2005), Research Paper 29, 31 pp.

[9] Y. Chen, P. Diaconis, S.P. Holmes, J.S. Liu, Sequential Monte Carlo methods for statistical analysis of tables, J. Amer. Statist. Assoc. 100 (2005) 109–120.
[10] W. Cuckler, J. Kahn, Hamiltonian cycles in Dirac graphs, Combinatorica 29 (2009) 299–326.
[11] W. Cuckler, J. Kahn, Entropy bounds for perfect matchings and Hamiltonian cycles, Combinatorica 29 (2009) 327–335.
[12] I.J. Good, J.F. Crook, The enumeration of arrays and a generalization related to contingency tables, Discrete Math. 19 (1977) 23–45.
[13] C. Greenhill, B.D. McKay, X. Wang, Asymptotic enumeration of sparse 0–1 matrices with irregular row and column sums, J. Combin. Theory Ser. A 113 (2006) 291–324.
[14] C. Greenhill, B.D. McKay, Random dense bipartite graphs and directed graphs with specified degrees, Random Structures Algorithms 35 (2009) 222–249.
[15] L. Gurvits, Classical complexity and quantum entanglement, J. Comput. System Sci. 69 (2004) 448–484.
[16] L. Gurvits, Van der Waerden/Schrijver–Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all, Electron. J. Combin. 15 (1) (2008), Research Paper 66.
[17] M. Jerrum, A. Sinclair, Fast uniform generation of regular graphs, Theoret. Comput. Sci. 73 (1990) 91–100.
[18] M. Jerrum, A. Sinclair, E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, J. ACM 51 (2004) 671–697.
[19] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover Publications, Inc., New York, 1957.
[20] N. Linial, A. Samorodnitsky, A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, Combinatorica 20 (2000) 545–568.
[21] I.G. Macdonald, Symmetric Functions and Hall Polynomials, with contributions by A. Zelevinsky, second edition, Oxford Math. Monogr., Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, 1995.
[22] C. McDiarmid, On the method of bounded differences, in: Surveys in Combinatorics, 1989, Norwich, 1989, in: London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 148–188.
[23] Yu. Nesterov, A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, SIAM Stud. Appl. Math., vol. 13, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.
[24] P.E. O’Neil, Asymptotics and random matrices with row-sum and column-sum restrictions, Bull. Amer. Math. Soc. 75 (1969) 1276–1282.
[25] J.H. van Lint, R.M. Wilson, A Course in Combinatorics, second edition, Cambridge University Press, Cambridge, 2001.