WHAT DOES A RANDOM CONTINGENCY TABLE LOOK LIKE?

Alexander Barvinok

November 2009

Abstract. Let $R=(r_1, \ldots, r_m)$ and $C=(c_1, \ldots, c_n)$ be positive integer vectors such that $r_1+\ldots+r_m=c_1+\ldots+c_n$. We consider the set $\Sigma(R,C)$ of non-negative $m \times n$ integer matrices (contingency tables) with row sums R and column sums C as a finite probability space with the uniform measure. We prove that a random table $D \in \Sigma(R,C)$ is close with high probability to a particular matrix ("typical table") Z defined as follows. We let $g(x)=(x+1)\ln(x+1)-x\ln x$ for $x \geq 0$ and let $g(X)=\sum_{i,j} g(x_{ij})$ for a non-negative matrix $X=(x_{ij})$. Then $g(X)$ is strictly concave and attains its maximum on the polytope of non-negative $m \times n$ matrices X with row sums R and column sums C at a unique point, which we call the typical table Z.

1. Introduction and the main result

(1.1) Random contingency tables. 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

$$\sum_{i=1}^m r_i = \sum_{j=1}^n c_j = N.$$

A *contingency table* with *margins* (R,C) is a non-negative integer matrix $D=(d_{ij})$ with row sums R and column sums C:

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

1991 *Mathematics Subject Classification.* 15A52, 05A16, 60C05, 15A36.

Key words and phrases. contingency table, random matrix, transportation polytope.

This research was partially supported by NSF Grants DMS 0400617 and DMS 0856640 and a United States - Israel BSF grant 2006377.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Let $\Sigma(R, C)$ be the set of all contingency tables with margins (R, C). As is well known, $\Sigma(R, C)$ is non-empty and finite. Let us consider $\Sigma(R, C)$ as a finite probability space endowed with the uniform probability measure. In this paper we address the following question:

Suppose that $D \in \Sigma(R, C)$ is chosen at random. What is D likely to look like?

The problem is interesting in its own right, but the main motivation comes from statistics; see [Go63], [DE85], [DG95] and references therein. A contingency table $D = (d_{ij})$ may represent certain statistical data (for example, d_{ij} may be the number of people in a certain sample having the i-th hair color and the j-th eye color). 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 [DE85]. To answer this question we need to know what a random table $D \in \Sigma(R, C)$ looks like. Considerable effort was invested in finding an efficient (polynomial time) algorithm to *sample* a random table $D \in \Sigma(R, C)$; see [DG95], [D+97], [C+06]. Despite a number of successes, such an algorithm is still at large in many interesting situations. In this paper, we do not discuss how to sample a random table but describe instead what it is likely to look like.

We prove that a random contingency table D is close in a certain sense to some particular non-negative $m \times n$ matrix Z, which we call the *typical table.*

(1.2) The typical table. Let $\mathcal{P}(R, C)$ be the set of all $m \times n$ non-negative matrices $X = (x_{ij})$ with row sums R and column sums C:

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

$$x_{ij} \geq 0 \quad \text{for all} \quad i, j.$$

Geometrically, $\mathcal{P}(R, C)$ is a convex polytope of dimension $(m-1)(n-1)$, known as the *transportation polytope.* Let

$$g(x) = (x+1)\ln(x+1) - x \ln x \quad \text{for} \quad x \geq 0$$

and let

$$g(X) = \sum_{i,j} g(x_{ij})$$

for a non-negative matrix $X = (x_{ij})$. One can easily check that g is strictly concave and hence achieves a unique maximum $Z = (z_{ij})$ on $\mathcal{P}(R, C)$. We call Z the *typical table* with margins (R, C). Since the objective function g is concave, Z can be computed efficiently, both in theory and in practice, by existing methods of convex optimization, cf. [NN94].

The solution Z to the above optimization problem was first introduced in the author's paper [Ba09]. It was given the name of "typical table" (perhaps with not enough justification) in [B+08].

In this paper, we show that Z indeed captures some typical features of a random table $D \in \Sigma(R, C)$.

We prove our main result assuming certain regularity ("smoothness") of margins.

(1.3) Smooth margins. Let us fix a number $0 < \delta \leq 1$. First, we assume that the row sums and column sums are of the same order:

$$\begin{aligned} \frac{\delta N}{m} \ \leq \ r_i \ \leq \ \frac{N}{\delta m} \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \\ \frac{\delta N}{n} \ \leq \ c_j \ \leq \ \frac{N}{\delta n} \quad \text{for} \quad j = 1, \ldots, n. \end{aligned} \tag{1.3.1}$$

Second, we assume that the *density* of the table is separated from 0:

$$\frac{N}{mn} \geq \delta. \tag{1.3.2}$$

We say that the margins (R, C) are δ*-smooth* if conditions (1.3.1)–(1.3.2) are satisfied. This is a modification of the definition from [B+08]. We note that δ-smooth margins are also δ'-smooth for any $0 < \delta' < \delta$. As we remarked (see (1.3.2)), we are interested in tables with the density separated from 0. For the case of sparse tables, where $r_i \ll n$ and $c_j \ll m$, see [Ne69], [GM08] and references therein.

Without loss of generality, we assume that $n \geq m$.

(1.4) Definitions and notation. Let us choose a non-empty subset of entries of a matrix:

$$S \subset \Big\{(i,j): \quad 1 \leq i \leq m, \quad 1 \leq j \leq n\Big\}.$$

For an $m \times n$ matrix $A = (a_{ij})$ let

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

be the sum of the entries from S.

The cardinality of a finite set X is denoted by $|X|$.

Now we state our main result.

(1.5) Theorem. *Let us fix real numbers $0 < \delta \leq 1$ and $\kappa > 0$. Then there exists a positive integer $q = q(\delta, \kappa)$ such that the following holds:*

Suppose that (R, C) are δ-smooth margins such that $n \geq m \geq q$.

Let

$$S \subset \{(i,j): \quad 1 \leq i \leq m, \quad 1 \leq j \leq n\}$$

be a set such that

$$|S| \ \geq \ \delta mn,$$

let Z be the typical table with margins (R, C), and let

$$\epsilon = \delta \frac{\ln n}{m^{1/3}}.$$

If $\epsilon \leq 1$ *then*

$$\mathbf{Pr}\Big\{D \in \Sigma(R,C) : \\ (1-\epsilon)\sigma_S(Z) \ \leq \sigma_S(D) \leq \ (1+\epsilon)\sigma_S(Z)\Big\} \ \geq \ 1 - 2n^{-\kappa n}.$$

In other words, asymptotically, as far as the sum over a positive fraction of entries is concerned, a contingency table D sampled uniformly at random from the set of contingency tables with given margins is very likely to be close to the typical table Z.

(1.6) The independence table. In [Go63], I.J. Good observes that the *independence table*

$$Y = (y_{ij}), \quad y_{ij} = r_i c_j / N \quad \text{for all} \quad i, j,$$

maximizes the entropy

$$H(X) = \sum_{i,j} \frac{x_{ij}}{N} \ln \frac{N}{x_{ij}}$$

on the set of all matrices $X = (x_{ij})$ in the transportation polytope $\mathcal{P}(R, C)$. One may be tempted to think that the independence table Y, not the typical table Z, reflects the structure of a random table $D \in \Sigma(R, C)$.

One can show that $Y = Z$ if and only if all row sums r_i are equal or all column sums c_j are equal. In fact, particular entries of the matrices Z and Y may demonstrate very different behavior even for reasonably looking margins. Suppose, for example, that $m = n$, that $r_1 = c_1 = 3n$ and that $r_i = c_i = n$ for $i > 1$. Hence $N = 3n + n(n-1) = n^2 + 2n$ and for the independence table we have

$$y_{11} = \frac{9n^2}{n^2 + 2n} \leq 9.$$

On the other hand, for the typical table Z the entry z_{11} grows linearly in n. Indeed, the optimality condition for Z (the gradient of g at Z is orthogonal to the affine span of the transportation polytope) implies that

$$\ln \left(\frac{z_{ij} + 1}{z_{ij}} \right) = \lambda_i + \mu_j \quad \text{for all} \quad i, j$$

and some $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n$; see Section 2.3. By symmetry, we can choose $\lambda_1 = \mu_1 = \alpha$ and $\lambda_i = \mu_i = \beta$ for $i > 1$. Moreover, we must have $0 < \alpha < \beta$. Since

$$z_{21} = \frac{1}{e^{\alpha+\beta} - 1} > \frac{1}{e^{2\beta} - 1} = z_{2j} \quad \text{for all} \quad j > 1$$

and $r_2 = n$, we should have

$$\beta > \frac{\ln 2}{2}.$$

Therefore,

$$z_{1j} = \frac{1}{e^{\alpha+\beta}-1} \ < \ \frac{1}{e^{\beta}-1} \ < \ \frac{1}{\sqrt{2}-1} \quad \text{for} \quad j > 1.$$

Since $r_1 = 3n$ we must have

$$z_{11} \ > \ 3n - \frac{n}{\sqrt{2}-1} > 0.58n.$$

Let us show that the independence table Y and the typical table Z may also produce different asymptotic behavior of the sums $\sigma_S(Y)$ and $\sigma_S(Z)$ as m and n grow and S is a subset of entries consisting of a positive fraction of all entries as in Theorem 1.5. For that, let us fix some margins $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ such that $z_{11} \neq y_{11}$. For a positive integer k let us consider the "cloned" margins

$$\begin{aligned} R_k &= \Big(\underbrace{kr_1, \ldots, kr_1}_{k \text{ times}}, \ldots, \underbrace{kr_m, \ldots, kr_m}_{k \text{ times}}\Big) \quad \text{and} \\ C_k &= \Big(\underbrace{kc_1, \ldots, kc_1}_{k \text{ times}}, \ldots, \underbrace{kc_n, \ldots, kr_n}_{k \text{ times}}\Big). \end{aligned} \tag{1.6.1}$$

In particular, tables $D \in \Sigma(R_k, C_k)$ are $km \times kn$ matrices whose total sum of entries is equal to $k^2 N$, where $N = r_1 + \ldots + r_m = c_1 + \ldots + c_n$. Let $S = S_k$ be the set of entries in the upper left $k \times k$ corner of a matrix from $\Sigma(R_k, C_k)$, let Y_k be the independence table of margins (R_k, C_k) and let Z_k be the typical table of margins (R_k, C_k). It is not hard to show that $\sigma_S(Z_k) = k^2 z_{11}$ and $\sigma_S(Y_k) = k^2 y_{11}$, so the ratio between the two sums remains fixed (and not equal to 1) as k grows.

It looks plausible that the independence table Y is indeed close with high probability to a random table $D \in \Sigma(R, C)$, if, instead of the uniform distribution in $\Sigma(R, C)$, a table $D = (d_{ij})$ is sampled from the Fisher-Yates probability measure, where

$$\mathbf{Pr}\,(D) = (N!)^{-1} \left(\prod_{i=1}^{m} r_i!\right) \left(\prod_{j=1}^{n} c_j!\right) \left(\prod_{ij} \frac{1}{d_{ij}!}\right);$$

see [DG95]. Compared with the uniform distribution, the Fisher-Yates measure gives less weight to tables with large entries.

Let $p, q > 0$ be real numbers such that $p + q = 1$. Recall that a discrete random variable x has *geometric distribution* if

$$\mathbf{Pr}\,\{x = k\} = pq^k \quad \text{for} \quad k = 0, 1, \ldots$$

We have

$$\mathbf{E}\, x = \frac{q}{p}.$$

Consequently,

$$\text{if} \quad \mathbf{E}\, x = z \quad \text{then} \quad p = \frac{1}{1+z} \quad \text{and} \quad q = \frac{z}{1+z}.$$

The following interpretation of the typical matrix was suggested to the author by J.A. Hartigan; see [BH09].

(1.7) Theorem. *Let $Z = (z_{ij})$ be the $m \times n$ typical table with margins (R, C). Let $X = (x_{ij})$ be the random $m \times n$ matrix of independent geometric random variables x_{ij} such that*

$$\mathbf{E}\, x_{ij} = z_{ij} \quad \text{for all} \quad i, j.$$

Then the probability mass function of X is constant on the set $\Sigma(R, C)$ of contingency tables with margins (R, C), and, moreover,

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

where g is the function defined in Section 1.2.

In other words, the multivariate geometric distribution X whose expectation is the typical matrix Z, when conditioned on the set $\Sigma(R, C)$ of contingency tables, results in the uniform probability distribution on $\Sigma(R, C)$. It turns out that for a positive $m \times n$ matrix A the value of $g(A)$ is equal to the maximum possible entropy of a random matrix with expectation A and values in the set $\mathbb{Z}_+^{m \times n}$ of $m \times n$ non-negative integer matrices. Such a maximum entropy random matrix is necessarily a matrix with independent geometrically distributed entries. Therefore, the distribution of X in Theorem 1.7 can be characterized as the maximum entropy distribution in the class consisting of all probability distributions on $\mathbb{Z}_+^{m \times n}$ whose expectations lie in the affine subspace consisting of the matrices with row sums R and column sums C; see [BH09].

(1.8) Possible ramifications and open questions. Theorem 1.7 allows one to interpret Theorem 1.5 as a law of large numbers for contingency tables: with respect to sums $\sigma_S(D)$ for sufficiently large sets S of entries, a random contingency table $D \in \Sigma(R, C)$ behaves approximately as the matrix of independent geometric variables whose expectation is the typical table. Similar concentration results can be obtained for other well-behaved functions on contingency tables. One can ask whether the distribution of a *particular entry* of a random table $D \in \Sigma(R, C)$ is asymptotically geometric, as the dimensions m and n of the table grow. For example, does the first entry d_{11} of the table converge in distribution to the geometric random variable with expectation z_{11} when the margins (R, C) are cloned, $(R, C) \longmapsto (R_k, C_k)$, as in (1.6.1)?

Let us fix a subset

$$W \subset \Big\{(i, j) : \quad i = 1, \ldots, m; \ j = 1, \ldots, n\Big\}.$$

Let us consider the set $\Sigma(R, C; W)$ of $m \times n$ non-negative integer matrices $D = (d_{ij})$ with row sums R, column sums C and such that $d_{ij} = 0$ for $(i, j) \notin W$. Assuming that $\Sigma(R, C; W)$ is non-empty, we can consider $\Sigma(R, C; W)$ as a finite probability space with the uniform measure and ask what a random table $D \in \Sigma(R, C; W)$ looks like.

As above, we define the typical table Z as the unique maximum of $g(X)$ on the polytope of non-negative matrices $X = (x_{ij})$ with row sums R, column sums C

and such that $x_{ij} = 0$ for $(i, j) \notin W$. One can prove versions of Theorem 1.5 and Theorem 1.7 in this more general context for subsets $S \subset W$. However, it appears that for Theorem 1.5 one has to assume, additionally, that there are no too large or too small values among the entries z_{ij} of the typical table $Z = (z_{ij})$, cf. the example in Section 1.6. In our case, when W is the set of all pairs (i, j), Lemma 2.4 ensures that the entries z_{ij} are not too small while Lemma 3.3 ensures that they are not too large.

In [Ba08] another variation of the problem is considered: what if we require $d_{ij} \in \{0, 1\}$ for all i, j. It turns out that a random D is close to a particular matrix maximizing the sum of entropies of the entries among all matrices with row sums R, column sums C and entries between 0 and 1.

In the rest of the paper, we prove Theorem 1.5.

In Section 2, we recall the main results of [Ba09] connecting the typical table Z with an asymptotic estimate for the number $|\Sigma(R, C)|$ of tables and also prove Theorem 1.7.

In Section 3, we prove Theorem 1.5 under the additional assumption that the total sum N of the entries is bounded by a polynomial in m and n.

In Section 4, we complete the proof of Theorem 1.5.

2. Preliminaries: an asymptotic formula for the number of tables

In [Ba09], the following result was proved; see Theorem 1.1 there.

(2.1) Theorem. *Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive integer vectors such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n = N$. Let us define a function*

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

Then $F(\mathbf{x}, \mathbf{y})$ attains its infimum

$$\rho(R, C) = \min_{\substack{0 < x_1, \ldots, x_m < 1 \\ 0 < y_1, \ldots, y_n < 1}} F(\mathbf{x}, \mathbf{y})$$

on the open cube $0 < x_i, y_j < 1$ and for the number $|\Sigma(R, C)|$ of non-negative integer $m \times n$ matrices with row sums R and column sums C we have

$$\rho(R, C) \ \geq \ |\Sigma(R, C)| \ \geq \ N^{-\gamma(m+n)} \rho(R, C),$$

where $\gamma > 0$ is an absolute constant.

□

As is remarked in [Ba09], the substitution $x_i = e^{-s_i}$, $y_j = e^{-t_j}$ transforms $\ln F(\mathbf{x}, \mathbf{y})$ into a convex function

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

on the positive orthant $\mathbb{R}_+^m \times \mathbb{R}_+^n$. It turns out that the typical table Z is the solution to the problem that is convex dual to the problem of minimizing G. The following result was proved in [Ba09]; see Lemma 1.4 there.

(2.2) Lemma. *Let $\mathcal{P} = \mathcal{P}(R, C)$ be the polytope of $m \times n$ non-negative matrices $X = (x_{ij})$ with row sums R and column sums C and let $Z \in \mathcal{P}(R, C)$ be the typical table; see Section 1.2.*

Then one can write $Z = (z_{ij})$,

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

and some $0 < \xi_1, \ldots, \xi_m; \eta_1, \ldots, \eta_n < 1$ such that the minimum $\rho(R, C)$ of the function $F(\mathbf{x}, \mathbf{y})$ in Theorem 2.1 is attained at $x^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$:

$$F(\mathbf{x}^*, \mathbf{y}^*) = \rho(R, C) = \min_{\substack{0 < x_1, \ldots, x_m < 1 \\ 0 < y_1, \ldots, y_n < 1}} F(\mathbf{x}, \mathbf{y}).$$

Moreover,

$$\rho(R, C) = \exp\left\{g(Z)\right\}.$$

□

Theorem 1.7 is a particular case of a more general result proved in [BH09]. Nevertheless, we present the proof of Theorem 1.7 here for completeness and since some elements of the proof will be recycled later.

(2.3) Proof of Theorem 1.7. From Lemma 2.2, we have $z_{ij} > 0$ for all i, j. Since Z lies in the relative interior of the transportation polytope $\mathcal{P}(R, C)$, the gradient of g at Z must be orthogonal to the subspace of $m \times n$ matrices with row and column sums equal to 0. Therefore,

$$\ln\left(\frac{z_{ij} + 1}{z_{ij}}\right) = \lambda_i + \mu_j \quad \text{for all} \quad i, j \tag{2.3.1}$$

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

For the geometric random variables x_{ij} we have

$$\mathbf{Pr}\left\{x_{ij}=d_{ij}\right\}=p_{ij}q_{ij}^{d_{ij}}=\left(\frac{1}{1+z_{ij}}\right)\left(\frac{z_{ij}}{1+z_{ij}}\right)^{d_{ij}}$$

Using (2.3.1), for $D\in\Sigma(R,C)$, $D=(d_{ij})$, we obtain

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

Also,

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

which completes the proof. □

We will need a lower bound for the entries of the typical table $Z=(z_{ij})$ proved in [B+08]; see Theorem 3.3 there.

(2.4) Lemma. *Let*

$$r_+=\max_{i=1,\dots,m} r_i,\quad r_-=\min_{i=1,\dots,m} r_i\quad\text{and}$$
$$c_+=\max_{j=1,\dots,n} c_j,\quad c_-=\min_{j=1,\dots,n} c_j.$$

Let $Z=(z_{ij})$ be the typical table with margins (R,C). Then

$$z_{ij}\ \geq\ \frac{r_-c_-}{r_+m}\quad\text{and}\quad z_{ij}\ \geq\ \frac{c_-r_-}{c_+n}\quad\text{for all}\quad i,j.$$

□

(2.5) Corollary. *Let $Z=(z_{ij})$ be the typical table of δ-smooth margins (R,C). Then*

$$z_{ij} \ \geq \ \frac{\delta^3 N}{mn} \quad \textit{for all} \quad i, j.$$

Proof. In Lemma 2.4, we have

$$r_- \ \geq \ \frac{\delta N}{m}, \quad c_- \ \geq \ \frac{\delta N}{n} \quad \text{and} \quad r_+ \ \leq \ \frac{N}{\delta m},$$

and the result follows. □

3. Proof of Theorem 1.5 assuming that N is polynomially bounded

In this section we prove Theorem 1.5 under the additional assumption that the total sum N of entries is bounded by a polynomial in m and n, specifically that $N \leq (mn)^{1/\delta}$. We use Theorem 1.7. We start with a standard large deviation inequality.

(3.1) Lemma. *Let $X=(x_{ij})$ be the $m \times n$ matrix of independent geometric random variables x_{ij} such that $\mathbf{E}\, X = Z$, $Z=(z_{ij})$. Let*

$$S \subset \Big\{ (i,j): \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \Big\}$$

be a non-empty set. Recall that

$$\sigma_S(X) = \sum_{(i,j) \in S} x_{ij}, \quad \sigma_S(Z) = \sum_{(i,j) \in S} z_{ij}$$

and let us denote

$$\nu_S(Z) = \sum_{(i,j) \in S} z_{ij}^2.$$

Then

(1) *For any real a and for any $0 < t \leq 2$, we have*

$$\mathbf{Pr}\left\{ \sigma_S(X) \ \leq \ -a + \sigma_S(Z) \right\} \ \leq \ \exp\left\{ -ta + \frac{t^2}{2}\big(\sigma_S(Z) + \nu_S(Z)\big) \right\}.$$

(2) *For any real a and for any $0 < t \leq \min\big\{1/3, \ 1/2z_{ij}: \ (i,j) \in S\big\}$, we have*

$$\mathbf{Pr}\left\{ \sigma_S(X) \ \geq \ a + \sigma_S(Z) \right\} \ \leq \ \exp\Big\{ -ta + 2t^2\big(\sigma_S(Z) + \nu_S(Z)\big) \Big\}.$$

Proof. We use the Laplace transform method; see, for example, Section 1.6 of [Le01]. To prove Part (1), for any $t > 0$ we compute

$$\mathbf{E}\, e^{-t\sigma_S(X)} = \prod_{(i,j)\in S} \mathbf{E}\, e^{-tx_{ij}} = \prod_{(i,j)\in S} \frac{p_{ij}}{1 - e^{-t} q_{ij}},$$

where

$$\mathbf{Pr}\left\{x_{ij} = k\right\} = p_{ij} q_{ij}^k \quad \text{for} \quad k = 0, 1, \ldots$$

Using the fact that $e^{-t} \leq 1 - t + t^2/2$ for $t \geq 0$, we obtain

$$\mathbf{E}\, e^{-t\sigma_S(X)} \leq \prod_{(i,j)\in S} \frac{p_{ij}}{p_{ij} + (t - t^2/2) q_{ij}} = \prod_{(i,j)\in S} \frac{1}{1 + (t - t^2/2) z_{ij}}.$$

Using the fact that $t - t^2/2 \geq 0$ for $0 \leq t \leq 2$ and that $\ln(1+x) \geq x - x^2/2$ for $x \geq 0$, we obtain

$$\begin{aligned}
\mathbf{E}\, e^{-t\sigma_S(X)} &\leq \exp\left\{ - \sum_{(i,j)\in S} \ln\Bigl(1 + \bigl(t - t^2/2\bigr) z_{ij}\Bigr) \right\} \\
&\leq \exp\left\{ - \sum_{(i,j)\in S} (t - t^2/2) z_{ij} + \frac{1}{2} \sum_{(i,j)\in S} (t - t^2/2)^2 z_{ij}^2 \right\} \\
&\leq \exp\left\{ -t\sigma_S(Z) + \frac{t^2}{2} \bigl(\sigma_S(Z) + \nu_S(Z)\bigr) \right\}.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{Pr}\left\{\sigma_S(X) \leq -a + \sigma_S(Z)\right\} &= \mathbf{Pr}\left\{-t\sigma_S(X) \geq ta - t\sigma_S(Z)\right\} \\
&= \mathbf{Pr}\left\{e^{-t\sigma_S(X)} \geq e^{ta - t\sigma_S(Z)}\right\} \\
&\leq e^{-ta + t\sigma_S(Z)} \mathbf{E}\, e^{-t\sigma_S(X)} \\
&\leq \exp\left\{ -ta + \frac{t^2}{2} \bigl(\sigma_S(Z) + \nu_S(Z)\bigr) \right\}.
\end{aligned}$$

To prove Part (2), we observe that $e^t < 1 + t + t^2$ for all $0 < t \leq 1$. Therefore, for $0 < t \leq \min\bigl\{1/3,\ 1/2z_{ij} :\ (i,j)\bigr\}$, we have

$$e^t < 1 + 2t \leq \frac{1 + z_{ij}}{z_{ij}} = \frac{1}{q_{ij}}$$

and hence

$$\begin{aligned}
\mathbf{E}\, e^{t\sigma_S(X)} &= \prod_{(i,j)\in S} \mathbf{E}\, e^{tx_{ij}} = \prod_{(i,j)\in S} \frac{p_{ij}}{1 - e^t q_{ij}} \\
&\leq \prod_{(i,j)\in S} \frac{p_{ij}}{p_{ij} - (t + t^2) q_{ij}} = \prod_{(i,j)\in S} \frac{1}{1 - (t + t^2) z_{ij}}.
\end{aligned}$$

Since $t \leq 1/3$ we have $t+t^2 \leq (4/3)t$ and hence $(t+t^2)z_{ij} \leq 2/3$. Using the fact that $\ln(1-x) \geq -x-x^2$ for $0 \leq x \leq 2/3$, we obtain

$$\begin{aligned} \mathbf{E}\, e^{t\sigma_S(X)} \leq & \exp\left\{ -\sum_{(i,j)\in S} \ln\left(1-(t+t^2)z_{ij}\right)\right\} \\ \leq & \exp\left\{ \sum_{(i,j)\in S} \left(t+t^2\right) z_{ij} + \sum_{(i,j)\in S} \left(t+t^2\right)^2 z_{ij}^2 \right\} \\ \leq & \exp\Big\{ t\sigma_S(Z) + 2t^2\left(\sigma_S(Z)+\nu_S(Z)\right)\Big\}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{Pr}\left\{\sigma_S(X) \ \geq \ a+\sigma_S(Z)\right\} &= \mathbf{Pr}\left\{t\sigma_S(X) \ \geq \ ta+t\sigma_S(Z)\right\} \\ &= \mathbf{Pr}\left\{e^{t\sigma_S(X)} \ \geq \ e^{ta+t\sigma_S(Z)}\right\} \\ &\leq e^{-ta-t\sigma_S(Z)}\mathbf{E}\, e^{t\sigma_S(X)} \\ &\leq \exp\Big\{-ta+2t^2\big(\sigma_S(Z)+\nu_S(Z)\big)\Big\}. \end{aligned}$$

□

One can observe that $\sigma_S(Z)+\nu_S(Z)$ is the variance of $\sigma_S(X)$.

(3.2) Corollary. *Let (R,C) be δ-smooth margins with the typical table $Z=(z_{ij})$ and let $X=(x_{ij})$ be the matrix of independent geometric variables such that $\mathbf{E}\, X = Z$. Suppose that*

$$z_{ij} \ \leq \ \frac{\alpha N}{mn} \quad \textit{for all} \quad (i,j)\in S$$

and some $\alpha \geq 1$. Then

(1) *For any $0<\epsilon<1$ we have*

$$\mathbf{Pr}\left\{\sigma_S(X) \ \leq \ (1-\epsilon)\sigma_S(Z)\right\} \ \leq \ \exp\left\{-\frac{\epsilon^2\delta^4|S|}{2+2\delta\alpha}\right\}.$$

(2) *For any $0<\epsilon<1$ we have*

$$\mathbf{Pr}\left\{\sigma_S(X) \ \geq \ (1+\epsilon)\sigma_S(Z)\right\} \ \leq \ \exp\left\{-\frac{\epsilon^2\delta^4|S|}{8+8\delta\alpha}\right\}.$$

Proof. Choosing

$$a=\epsilon\sigma_S(Z) \quad \text{and} \quad t=\frac{\epsilon\sigma_S(Z)}{\sigma_S(Z)+\nu_S(Z)}$$

in Part (1) of Lemma 3.1, we obtain

$$\mathbf{Pr}\left\{\sigma_S(X) \ \leq \ (1-\epsilon)\sigma_S(Z)\right\} \ \leq \ \exp\left\{-\frac{\epsilon^2\sigma_S^2(Z)}{2(\sigma_S(Z)+\nu_S(Z))}\right\}. \tag{3.2.1}$$

Furthermore,

$$\nu_S(Z) = \sum_{(i,j)\in S} z_{ij}^2 \ \leq \ \frac{\alpha N}{mn}\sum_{(i,j)\in S} z_{ij} \ = \ \frac{\alpha N}{mn}\sigma_S(Z). \tag{3.2.2}$$

By Corollary 2.5,

$$\delta_S(Z) \ \geq \ |S|\frac{\delta^3 N}{mn}. \tag{3.2.3}$$

We recall that

$$\frac{N}{mn} \ \geq \ \delta. \tag{3.2.4}$$

Summarizing (3.2.1)–(3.2.4), we get

$$\begin{aligned}
\mathbf{Pr}\left\{\sigma_S(X) \ \leq \ (1-\epsilon)\sigma_S(Z)\right\} \ &\leq \ \exp\left\{-\frac{\epsilon^2\sigma_S(Z)mn}{2\,(mn+\alpha N)}\right\} \\
&\leq \ \exp\left\{-\frac{\epsilon^2|S|\delta^3 N}{2\,(mn+\alpha N)}\right\} \\
&\leq \ \exp\left\{-\frac{\epsilon^2|S|\delta^3}{2\,(mn/N+\alpha)}\right\} \\
&\leq \ \exp\left\{-\frac{\epsilon^2\delta^4|S|}{2+2\delta\alpha}\right\}
\end{aligned}$$

and Part (1) follows.

Let us choose $a=\epsilon\sigma_S(Z)$ in Part (2) of Lemma 3.1. Let

$$t_0 = \frac{\epsilon\sigma_S(Z)}{4\,(\sigma_S(Z)+\nu_S(Z))}.$$

Clearly, $t_0 \leq 1/4 < 1/3$. If $t_0 < mn/2\alpha N$, we choose $t=t_0$ and if $t_0 \geq mn/2\alpha N$, we choose $t = mn/2\alpha N$ in Part (2) of Lemma 3.1. Hence if $t_0 < mn/2\alpha N$, we obtain as above in Part (1)

$$\begin{aligned}
\mathbf{Pr}\left\{\sigma_S(X) \ \geq \ (1+\epsilon)\sigma_S(Z)\right\} \ &\leq \ \exp\left\{-\frac{\epsilon^2\sigma_S^2(Z)}{8\,(\sigma_S(Z)+\nu_S(Z))}\right\} \\
&\leq \ \exp\left\{-\frac{\epsilon^2\delta^4|S|}{8+8\delta\alpha}\right\}.
\end{aligned} \tag{3.2.5}$$

If $t_0 \geq mn/2\alpha N$ then

$$\sigma_S(Z) + \nu_S(Z) \ \leq \ \frac{\epsilon \sigma_S(Z) \alpha N}{2mn}.$$

Therefore, choosing $t = mn/2\alpha N$ in Part (2) of Lemma 3.1, we obtain

$$\begin{aligned} \mathbf{Pr}\left\{\sigma_S(X) \ \geq \ (1+\epsilon)\sigma_S(Z)\right\} \ &\leq \ \exp\left\{-\frac{\epsilon mn \sigma_S(Z)}{2\alpha N} + \frac{m^2 n^2 \left(\sigma_S(Z) + \nu_S(Z)\right)}{2\alpha^2 N^2}\right\} \\ &\leq \ \exp\left\{-\frac{\epsilon \sigma_S(Z) mn}{4\alpha N}\right\}. \end{aligned}$$

Using (3.2.3), we obtain

$$\mathbf{Pr}\left\{\sigma_S(X) \ \geq \ (1+\epsilon)\sigma_S(Z)\right\} \ \leq \ \exp\left\{-\frac{\epsilon \delta^3 |S|}{4\alpha}\right\}. \tag{3.2.6}$$

Comparing (3.2.5) and (3.2.6), we complete the proof. □

Now we can prove the following weaker version of Theorem 1.5.

(3.3) Proposition. *Let us fix real numbers $0 < \delta \leq 1$ and $\kappa > 0$. Then there exists a positive integer $q = q(\delta, \kappa)$ such that the following holds:*

Suppose that (R, C) are δ-smooth margins such that $n \geq m \geq q$ and let $Z = (z_{ij})$ be the typical table with margins (R, C). Let

$$S \subset \left\{(i,j): \quad 1 \leq i \leq m, \quad 1 \leq j \leq n\right\}$$

be a set such that

$$|S| \geq \delta mn$$

and suppose that the entries z_{ij} of the typical table satisfy the inequalities

$$z_{ij} \ \leq \ \frac{\alpha N}{mn} \quad \text{for} \quad \alpha = 2\delta^{-1} m^{1/3}$$

and all $(i,j) \in S$.

Suppose further that for the total sum N of entries we have

$$N \ \leq \ (mn)^{1/\delta}.$$

Let

$$\epsilon = \frac{\delta \ln n}{m^{1/3}}.$$

If $\epsilon \leq 1$, we have

$$\begin{aligned} &\mathbf{Pr}\left\{D \in \Sigma(R,C): \ \sigma_S(D) \ \leq \ (1-\epsilon)\sigma_S(Z)\right\} \ \leq \ n^{-\kappa n} \quad \text{and} \\ &\mathbf{Pr}\left\{D \in \Sigma(R,C): \ \sigma_S(D) \ \geq \ (1+\epsilon)\sigma_S(Z)\right\} \ \leq \ n^{-\kappa n}. \end{aligned}$$

Proof. Let $X = (x_{ij})$ be the $m \times n$ matrix of independent geometric random variables x_{ij} such that $\mathbf{E}\, X = Z$. By Theorem 1.7, the distribution of X conditioned on $X \in \Sigma(R,C)$ is uniform and hence

$$\begin{aligned}\mathbf{Pr}\left\{D \in \Sigma(R,C):\ \sigma_S(D) \leq (1-\epsilon)\sigma_S(Z)\right\} \\ = \frac{\mathbf{Pr}\left\{X:\ \sigma_S(X) \leq (1-\epsilon)\sigma_S(Z) \quad \text{and} \quad X \in \Sigma(R,C)\right\}}{\mathbf{Pr}\left\{X:\ X \in \Sigma(R,C)\right\}}.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbf{Pr}\left\{D \in \Sigma(R,C):\ \sigma_S(D) \geq (1+\epsilon)\sigma_S(Z)\right\} \\ = \frac{\mathbf{Pr}\left\{X:\ \sigma_S(X) \geq (1+\epsilon)\sigma_S(Z) \quad \text{and} \quad X \in \Sigma(R,C)\right\}}{\mathbf{Pr}\left\{X:\ X \in \Sigma(R,C)\right\}}.\end{aligned}$$

By Theorem 1.7, Lemma 2.2 and Theorem 2.1 we get

$$\mathbf{Pr}\left\{X \in \Sigma(R,C)\right\} \ = \ e^{-g(Z)}\left|\Sigma(R,C)\right| \ \geq \ N^{-\gamma(m+n)}$$

for some absolute constant $\gamma > 0$. Since $N \leq (mn)^{1/\delta}$, we obtain

$$\begin{aligned}\mathbf{Pr}\left\{D \in \Sigma(R,C):\ \sigma_S(D) \leq (1-\epsilon)\sigma_S(Z)\right\} \\ \leq \ (mn)^{\gamma_1(m+n)}\mathbf{Pr}\left\{X:\ \sigma_S(X) \leq (1-\epsilon)\sigma_S(Z)\right\}\end{aligned}$$

and similarly

$$\begin{aligned}\mathbf{Pr}\left\{D \in \Sigma(R,C):\ \sigma_S(D) \geq (1+\epsilon)\sigma_S(Z)\right\} \\ \leq \ (mn)^{\gamma_1(m+n)}\mathbf{Pr}\left\{X:\ \sigma_S(X) \geq (1+\epsilon)\sigma_S(Z)\right\}\end{aligned}$$

for some constant $\gamma_1 = \gamma(\delta) > 0$. By Part (1) of Corollary 3.2,

$$\mathbf{Pr}\left\{X:\ \sigma_S(X) \leq (1-\epsilon)\sigma_S(Z)\right\} \ \leq \ \exp\left\{-\frac{\delta^7 mn \ln^2 n}{m^{2/3}(2+4m^{1/3})}\right\},$$

while by Part (2) of Corollary 3.2

$$\mathbf{Pr}\left\{X:\ \sigma_S(X) \geq (1+\epsilon)\sigma_S(Z)\right\} \ \leq \ \exp\left\{-\frac{\delta^7 mn \ln^2 n}{m^{2/3}(8+16m^{1/3})}\right\},$$

and the result follows. □

Next, we prove that large entries of the typical table Z belong to a small number of rows.

(3.4) Lemma. *Let (R, C) be δ-smooth margins and let $Z = (z_{ij})$ be the $m \times n$ typical table with margins (R, C). Let $\alpha \geq 2mn/N$ be a real number. Let*

$$I = \left\{ i : \quad z_{ij} \geq \frac{\alpha N}{mn} \quad \text{for some} \quad j \right\}.$$

Then

$$|I| \leq \frac{4m}{\delta\alpha}.$$

Proof. By (2.3.1), we can write

$$\ln\left(\frac{z_{ij}+1}{z_{ij}}\right) = \lambda_i + \mu_j \quad \text{for all} \quad i, j$$

and some $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$. Since $\lambda_i + \mu_j > 0$ for all i and j, without loss of generality we may assume that $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$ are positive.

Let

$$I_0 = \left\{ i : \quad \lambda_i \leq \frac{mn}{\alpha N} \right\} \quad \text{and} \quad J_0 = \left\{ j : \quad \mu_j \leq \frac{mn}{\alpha N} \right\}.$$

If $i \in I$ then for some j we have

$$\frac{mn}{\alpha N} \geq \frac{1}{z_{ij}} \geq \ln\left(\frac{z_{ij}+1}{z_{ij}}\right) \geq \lambda_i$$

and therefore $I \subset I_0$. Similarly, if $z_{ij} \geq \alpha N/mn$ for some i then $j \in J_0$. Hence without loss of generality, we may assume that $J_0 \neq \emptyset$.

Let us fix a $j_0 \in J_0$. Then for any $i \in I_0$ we have

$$\ln\left(\frac{z_{ij_0}+1}{z_{ij_0}}\right) \leq \frac{2mn}{\alpha N}.$$

Hence for all $i \in I_0$ we have

$$\frac{1}{z_{ij_0}} \leq \exp\left\{\frac{2mn}{\alpha N}\right\} - 1 \leq \frac{4mn}{\alpha N}$$

(using the fact that $e^x \leq 1 + 2x$ for $0 \leq x \leq 1$). Hence

$$z_{ij_0} \geq \frac{\alpha N}{4mn} \quad \text{for} \quad i \in I_0.$$

Since

$$\sum_{i=1}^{m} z_{ij_0} = c_{j_0} \leq \frac{N}{\delta n},$$

we conclude that

$$|I| \leq |I_0| \leq \frac{4c_{j_0} mn}{\alpha N} \leq \frac{4m}{\delta\alpha}.$$

□

Finally, we prove the main result of this section.

(3.5) Proposition. *In Theorem 1.5 assume, additionally, that* $N \leq (mn)^{1/\delta}$ *(equivalently, drop the upper bound assumption for* z_{ij} *in Proposition 3.3). Then the conclusion of Theorem 1.5 holds (equivalently, the conclusion of Proposition 3.3 holds).*

Proof. Let us choose

$$\alpha = 2\delta^{-1} m^{1/3}$$

and let

$$I = \left\{ i : \quad z_{ij} \geq \frac{\alpha N}{mn} \quad \text{for some} \quad j \right\}.$$

Since $N/mn \geq \delta$, we have $\alpha \geq 2mn/N$ and by Lemma 3.4 we have

$$|I| \ \leq \ 2m^{2/3}.$$

Let

$$S_0 = \big\{ (i,j) \in S : \quad i \notin I \big\}.$$

Then

$$|S \setminus S_0| \ \leq \ n|I| \ \leq \ 2nm^{2/3},$$

and hence for $\delta_0 = \delta/2$ and m sufficiently large, $n \geq m \geq q(\delta)$, we have

$$|S_0| \ \geq \ \delta_0 mn.$$

Furthermore, we have

$$\sigma_{S \setminus S_0}(D), \ \sigma_{S \setminus S_0}(Z) \ \leq \ \sum_{i \in I} r_i \ \leq \ |I| \frac{N}{\delta m} \ \leq \ \frac{2N}{\delta m^{1/3}}.$$

On the other hand, by Corollary 2.5, we have

$$\sigma_S(Z) \ \geq \ |S| \frac{\delta^3 N}{mn} \ \geq \ \delta^4 N.$$

Therefore,

$$\sigma_{S_0}(Z) \ = \ \sigma_S(Z) - \sigma_{S \setminus S_0}(Z) \ \geq \ \left(1 - \frac{2}{\delta^5 m^{1/3}} \right) \sigma_S(Z) \tag{3.5.1}$$

and, similarly,

$$\sigma_S(Z) - \sigma_{S \setminus S_0}(D) \ \geq \ \left(1 - \frac{2}{\delta^5 m^{1/3}} \right) \sigma_S(Z). \tag{3.5.2}$$

We have

$$\begin{aligned}&\mathbf{Pr}\left\{D \in \Sigma(R,C):\ \sigma_S(D) \leq (1-\epsilon)\sigma_S(Z)\right\}\\ &\qquad \leq\ \mathbf{Pr}\left\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \leq (1-\epsilon)\sigma_S(Z)\right\}.\end{aligned}$$

By (3.5.1) we obtain

$$\begin{aligned}(1-\epsilon)\sigma_S(Z)\ &=\ \left(1-\frac{\delta \ln n}{m^{1/3}}\right)\sigma_S(Z)\ \leq\ (1-\epsilon_0)\left(1-\frac{2}{\delta^5 m^{1/3}}\right)\sigma_S(Z)\\ &\leq\ (1-\epsilon_0)\,\sigma_{S_0}(Z), \quad \text{where}\\ &\qquad\quad \epsilon_0=\frac{\delta \ln n}{2m^{1/3}},\end{aligned}$$

and m is sufficiently large, $n \geq m \geq q(\delta)$.

Applying Proposition 3.3 with $S_0 \subset S$ and $\delta_0 = \delta/2$, we conclude that if m is sufficiently large, $n \geq m \geq q(\delta, \kappa)$, we have

$$\begin{aligned}&\mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \leq (1-\epsilon)\sigma_S(Z)\big\}\\ &\qquad \leq\ \mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \leq (1-\epsilon_0)\,\sigma_{S_0}(Z)\big\}\ \leq\ n^{-\kappa n}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}&\mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_S(D) \geq (1+\epsilon)\sigma_S(Z)\big\}\\ &\qquad =\ \mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \geq (1+\epsilon)\sigma_S(Z) - \sigma_{S\setminus S_0}(D)\big\}\\ &\qquad \leq\ \mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \geq (1+\epsilon)\big(\sigma_S(Z) - \sigma_{S\setminus S_0}(D)\big)\big\}.\end{aligned}$$

By (3.5.2) we obtain

$$\begin{aligned}(1+\epsilon)\big(\sigma_S(Z) - \sigma_{S\setminus S_0}(D)\big)\ &\geq (1+\epsilon)\left(1-\frac{2}{\delta^5 m^{1/3}}\right)\sigma_S(Z)\\ &\geq\ (1+\epsilon_0)\,\sigma_{S_0}(Z), \quad \text{where}\\ &\qquad \epsilon_0=\frac{\delta \ln n}{2m^{1/3}},\end{aligned}$$

and m is sufficiently large, $n \geq m \geq q(\delta)$.

Applying Proposition 3.3 with $S_0 \subset S$ and $\delta_0 = \delta/2$, we conclude that if m is sufficiently large, $n \geq m \geq q(\delta, \kappa)$, we have

$$\begin{aligned}&\mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \geq (1+\epsilon)\,\big(\sigma_S(Z) - \sigma_{S\setminus S_0}(D)\big)\big\}\\ &\qquad \leq\ \mathbf{Pr}\big\{D \in \Sigma(R,C):\ \sigma_{S_0}(D) \geq (1+\epsilon_0)\,\sigma_{S_0}(Z)\big\}\ \leq\ n^{-\kappa}\end{aligned}$$

and the result follows. □

4. Proof of Theorem 1.5

It remains to prove Theorem 1.5 in the case of a large (superpolynomial in mn) total sum N of entries. More precisely, we assume that $N > (mn)^7$ since the case of $N \leq (mn)^7$ is covered by Proposition 3.5 with a sufficiently small $\delta \leq 1/7$ (we recall that δ-smooth margins are also δ'-smooth with any $0 < \delta' < \delta$).

The idea of the proof is as follows: given margins (R, C) whose total sum of entries is N, we construct new margins (R', C') whose total sum of entries N' is bounded by a polynomial in mn and a scaling map

$$\mathcal{T}: \quad \Sigma(R, C) \longrightarrow \Sigma(R', C'),$$

which, roughly, scales every table $D \in \Sigma(R, C)$ by the same factor t. We then deduce Theorem 1.5 for margins (R, C) from that for margins (R', C').

We have

$$R' \approx t^{-1} R, \quad C' \approx t^{-1} C \quad \text{and} \quad \mathcal{T}(D) \approx t^{-1} D,$$

where "$\approx$" stands for rounding in some consistent way.

In constructing the map $\mathcal{T}$ we essentially follow the ideas of [D+97].

(4.1) Lattices, bases, and fundamental parallelepipeds. Let $\mathcal{V}$ be a finite-dimensional real vector space and let $\Lambda \subset \mathcal{V}$ be a *lattice*, that is, a discrete additive subgroup of $\mathcal{V}$ which spans $\mathcal{V}$. Suppose that $\dim \mathcal{V} = k$ and let $u_1, \ldots, u_k$ be a basis of Λ. The set

$$\Pi = \left\{ \sum_{i=1}^{k} \lambda_i u_i : \quad 0 \leq \lambda_i < 1 \quad \text{for} \quad i = 1, \ldots, k \right\}$$

is called the *fundamental parallelepiped* associated with the basis $u_1, \ldots, u_k$.

Suppose that $\mathcal{A}$ is an affine space, with $\dim \mathcal{A} = \dim \mathcal{V}$, on which $\mathcal{V}$ acts by translations: $a + v \in \mathcal{A}$ for all $a \in \mathcal{A}$ and $v \in \mathcal{V}$ and $a + (v_1 + v_2) = (a + v_1) + v_2$ for all $a \in \mathcal{A}$ and $v_1, v_2 \in \mathcal{V}$. Let us choose $a \in \mathcal{A}$. The set $\Lambda_a = a + \Lambda$ is called a *point lattice* in $\mathcal{A}$. As is known, the translations $v + \Pi : \ v \in \Lambda_a$ cover $\mathcal{A}$ without overlapping.

We will also use the following standard fact. Suppose that $\Lambda_1 \supset \Lambda$ is a finer lattice and let $|\Lambda_1/\Lambda| < \infty$ be its index. Then, for any $a, b \in \mathcal{A}$ we have

$$|(a + \Pi) \cap (b + \Lambda_1)| = |\Lambda_1/\Lambda|,$$

see for example Chapter VII of [Ba02].

Let us fix a point lattice $\Lambda_a \subset \mathcal{A}$ and a fundamental parallelepiped $\Pi \subset \mathcal{V}$ of Λ. Given a point $x \in \mathcal{A}$, we define its *rounding* $y = \lfloor x \rfloor_{\Lambda_a, \Pi}$ as the unique point $y \in \Lambda_a$ such that $x \in y + \Pi$.

In our case, $\mathcal{V}$ is the space of real $m \times n$ matrices with the row and column sums equal to 0, so $\dim \mathcal{V} = (m-1)(n-1)$, while $\mathcal{A}$ is the affine space of $m \times n$ matrices with prescribed integer row and column sums, so that for all $D \in \mathcal{A}$ and $U \in \mathcal{V}$ we have $D + U \in \mathcal{A}$. Furthermore, let $\Lambda \subset \mathcal{V}$ be the lattice of integer matrices and let $\Lambda' \subset \mathcal{A}$ be the point lattice consisting of integer matrices.

As is shown, for example, in [D+97], lattice Λ has a basis consisting of the matrices U_{ij} for $1 \leq i \leq n-1$, $1 \leq j \leq m-1$ that have 1 in the (i, j) and $(i+1, j+1)$ positions, -1 in the $(i+1, j)$ and $(i, j+1)$ positions and zeros elsewhere. Let Π be the fundamental parallelepiped of this basis $\{U_{ij}\}$. We call this parallelepiped Π *standard*. We note that

$$-2 \leq x_{ij} \leq 2 \quad \text{for all} \quad i, j \quad \text{and all} \quad X \in \Pi, \quad X = (x_{ij}). \tag{4.1.1}$$

Finally, for positive integer t let $\Lambda_1 = t^{-1}\Lambda$. Hence $|\Lambda_1/\Lambda| = t^{(m-1)(n-1)}$.

(4.2) The t-scaling map $\mathcal{T}$. Let us choose a positive integer t and an arbitrary $D_0 \in \Sigma(R, C)$, where $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$. Let us define a positive $m \times n$ matrix B as follows. First, we obtain D_1 by rounding up to the nearest integer every entry of $t^{-1}D_0$ and adding 2 to the result. In particular, D_1 is a positive integer matrix. Let

$$B = D_1 - t^{-1}D_0, \quad \text{so} \quad D_1 = B + t^{-1}D_0.$$

Clearly, $B = (b_{ij})$ is an $m \times n$ matrix with

$$2 \leq b_{ij} < 3 \quad \text{for all} \quad i, j. \tag{4.2.1}$$

Let $R' = (r'_1, \ldots, r'_m)$ and $C' = (c'_1, \ldots, c'_n)$ be the row and column sums of D_1 respectively. Thus R' and C' are positive integer vectors and

$$\begin{gathered} t^{-1}r_i + 2n \ \leq \ r'_i \ \leq \ t^{-1}r_i + 3n \quad \text{for} \quad i = 1, \ldots, m \\ \text{and} \\ t^{-1}c_j + 2m \ \leq \ c'_j \ \leq \ t^{-1}c_j + 3m \quad \text{for} \quad j = 1, \ldots, n. \end{gathered} \tag{4.2.2}$$

Let $\mathcal{A}$ be the affine subspace of matrices with row sums R' and column sums C' and let $\Lambda' \subset \mathcal{A}$ be the point lattice of integer matrices. Thus $\Lambda' = D_1 + \Lambda$, where Λ is the lattice of $m \times n$ integer matrices with zero row and column sums, see Section 4.1. For a matrix $D \in \Sigma(R, C)$ we define a matrix $\mathcal{T}(D)$ by

$$\mathcal{T}(D) = \lfloor t^{-1}D + B \rfloor_{\Lambda', \Pi},$$

where Π is the standard parallelepiped of Λ; see Section 4.1. In words: given a table $D \in \Sigma(R, C)$, matrix $\mathcal{T}(D)$ is the unique integer matrix such that the translation $\mathcal{T}(D) + \Pi$ of the standard parallelepiped Π contains $t^{-1}D + B$. Clearly, $\mathcal{T}(D)$ is an $m \times n$ integer matrix with row sums R' and column sums C'. Moreover, since every entry of $t^{-1}D + B$ is at least 2 and because of (4.1.1), matrix $\mathcal{T}(D)$ is non-negative.

Hence we have defined a map

$$\mathcal{T} : \Sigma(R, C) \longrightarrow \Sigma(R', C').$$

We summarize some of its properties below.

(4.3) Lemma.

(1) *For all* $Y \in \Sigma(R', C')$ *we have*

$$|\mathcal{T}^{-1}(Y)| \leq t^{(m-1)(n-1)};$$

(2) *Let* $S \subset \Big\{(i,j): \ i=1, \ldots, m, \ j=1, \ldots, n\Big\}$ *be a set of indices. Then*

$$t^{-1}\sigma_S(D) \ \leq \ \sigma_S(\mathcal{T}(D)) \ \leq \ t^{-1}\sigma_S(D) + 5|S|$$

for all $D \in \Sigma(R, C)$.

Proof. Given $Y \in \Sigma(R', C')$, we compute $\mathcal{T}^{-1}(Y)$ as follows: we consider the translation $(Y - B) + \Pi$ of the standard parallelepiped Π and observe that

$$\mathcal{T}^{-1}(Y) = \Big\{D: \quad t^{-1}D \in (Y-B) + \Pi \quad \text{and} \\ D \quad \text{is a non-negative integer matrix}\Big\}.$$

Recall that $\Lambda \subset \mathcal{V}$ is the lattice of $m \times n$ integer matrices with the row and column sums equal to 0 and that $\Lambda_1 = t^{-1}\Lambda$. In the affine space of $m \times n$ matrices with row sums $t^{-1}R$ and column sums $t^{-1}C$ let us consider the point lattice $\Lambda_1' = t^{-1}D_0 + \Lambda_1$ consisting of matrices $t^{-1}D$ where D is an integer matrix. Then

$$|((Y-B)+\Pi) \cap \Lambda_1'| = |\Lambda_1 / \Lambda| = t^{(m-1)(n-1)}$$

and Part (1) follows. Part (2) follows because of (4.1.1) and (4.2.1). $\square$

(4.4) Lemma. *Suppose that*

$$r_i', \ c_j' \ \geq \ (mn)^2 \quad \textit{for all} \quad i, j.$$

Then, for any $\zeta \geq 0$ *we have*

$$\mathbf{Pr}\Big\{D \in \Sigma(R,C): \ \sigma_S(D) \geq t\zeta\Big\} \ \leq \ \beta \mathbf{Pr}\Big\{Y \in \Sigma(R', C'): \ \sigma_S(Y) \geq \zeta\Big\}$$

and

$$\mathbf{Pr}\Big\{D \in \Sigma(R,C): \ \sigma_S(D) \leq t\zeta\Big\} \ \leq \ \beta \mathbf{Pr}\Big\{Y \in \Sigma(R', C'): \ \sigma_S(Y) \leq \zeta + 5|S|\Big\},$$

where $\beta > 0$ *is an absolute constant.*

Proof. By Part (2) of Lemma 4.3, if $\sigma_S(D) \geq t\zeta$ then $\sigma_S(Y) \geq \zeta$ for $Y = \mathcal{T}(D)$. Using Part (1) of Lemma 4.3, we can write

$$\begin{aligned}\mathbf{Pr}\Big\{D \in \Sigma(R,C): \sigma_S(D) \geq t\zeta\Big\} =& \frac{\Big|D \in \Sigma(R,C): \ \sigma_S(D) \geq t\zeta\Big|}{|\Sigma(R,C)|} \\ &\leq t^{(m-1)(n-1)} \frac{\Big|Y \in \Sigma(R',C'): \ \sigma_S(Y) \geq \zeta\Big|}{|\Sigma(R,C)|} \\ &= \frac{|\Sigma(R',C')|}{|\Sigma(R,C)|} t^{(m-1)(n-1)} \mathbf{Pr}\Big\{Y \in \Sigma(R',C'): \ \sigma_S(Y) \geq \zeta\Big\}.\end{aligned}$$

Similarly, by Part (2) of Lemma 4.3, if $\sigma_S(D) \leq t\zeta$ then $\sigma_S(Y) \leq \zeta + 5|S|$ for $Y = \mathcal{T}(D)$ and

$$\begin{aligned}\mathbf{Pr}\Big\{D \in \Sigma(R,C): \ \sigma_S(D) \leq t\zeta\Big\} \\ \leq \frac{|\Sigma(R',C')|}{|\Sigma(R,C)|} t^{(m-1)(n-1)} \mathbf{Pr}\Big\{Y \in \Sigma(R',C'): \ \sigma_S(Y) \leq \zeta + 5|S|\Big\}.\end{aligned}$$

It is shown in [D+97] that for sufficiently large margins, the number of contingency tables is approximated within a constant factor by the volume of the corresponding transportation polytope; see Section 1.2. In particular, estimates of [D+97] imply that

$$|\Sigma(R',C')| \ \leq \ \beta_1 \operatorname{vol} \mathcal{P}(R',C') \quad \text{and} \quad |\Sigma(R,C)| \ \geq \ \beta_2 \operatorname{vol} \mathcal{P}(R,C)$$

for some absolute constants $\beta_1, \beta_2 > 0$.

From (4.2.2), we have

$$\begin{aligned} r_i \ &\geq \ t(r_i' - 3n) \ \geq \ tr_i'\left(1 - \frac{3}{m^2 n}\right) \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \\ c_j \ &\geq \ t(r_i' - 3m) \ \geq \ tc_j'\left(1 - \frac{3}{mn^2}\right) \quad \text{for} \quad j = 1, \ldots, n. \end{aligned}$$

It follows then that

$$\operatorname{vol} \mathcal{P}(R,C) \ \geq \ \beta_3 t^{(m-1)(n-1)} \operatorname{vol} \mathcal{P}(R',C')$$

for some absolute constant $\beta_3 > 0$. The result now follows. □

Next, we show that the t-scaling map $\mathcal{T}$ almost scales the typical table provided the margins R', C' are large enough, that is, $Z' \approx t^{-1}Z$. The idea of the proof is roughly the following: if margins (R', C') and (R, C) are large enough, then the corresponding typical tables Z' and Z roughly optimize the functional $\sum_{i,j} \ln x_{ij}$ on the corresponding transportation polytopes and hence the map $X \longmapsto tX$ roughly maps Z' to Z.

(4.5) Lemma. *Let $Z = (z_{ij})$ be the typical table with margins (R, C), let $Z' = (z_{ij}')$ be the typical table with margins (R', C') obtained by t-scaling and suppose that*

$$z_{ij}' \ \geq \ (mn)^4 + 3 \quad \text{for all} \quad i, j.$$

Then

$$\left|\frac{z_{ij}}{tz_{ij}'} - 1\right| \ \leq \ \frac{\beta}{mn} \quad \text{for all} \quad i, j$$

and some absolute constant $\beta > 0$.

Proof. First, we prove some useful inequalities for the function

$$g(x) = (x+1)\ln(x+1) - x\ln x.$$

We have

$$g(tx) - g(x) = \int_x^{tx} g'(y)\, dy = \int_x^{tx} \ln\left(\frac{y+1}{y}\right)\, dy \leq \int_x^{tx} \frac{dy}{y} = \ln(tx) - \ln x = \ln t.$$

Also,

$$\begin{aligned} g(x) =&(x+1)\ln(x+1) - (x+1)\ln x + (x+1)\ln x - x\ln x \\ =&(x+1)\ln\left(\frac{x+1}{x}\right) + \ln x = \ln x + 1 + O\left(\frac{1}{x}\right) \quad \text{for} \quad x \geq 1. \end{aligned}$$

Finally, we note that

$$g''(x) = -\frac{1}{x(x+1)}.$$

Since from (4.2.2) we have

$$r_i \ \leq \ tr_i' \quad \text{and} \quad c_j \ \leq \ tc_j' \quad \text{for all} \quad i,j$$

we have

$$\max_{X \in \mathcal{P}(R,C)} g(X) \ \leq \ \ln t + \max_{X \in \mathcal{P}(R',C')} g(X). \tag{4.5.1}$$

Let B be the matrix constructed in Section 4.2 and let $W = t(Z' - B) \in \mathcal{P}(R,C)$. Hence

$$w_{ij} \ \geq \ t(mn)^4 \quad \text{for all} \quad i,j.$$

Since

$$g\left(w_{ij}\right) = 1 + \ln w_{ij} + O\left(\frac{1}{m^4n^4}\right) \quad \text{and} \quad g\left(z_{ij}'\right) = 1 + \ln z_{ij}' + O\left(\frac{1}{m^4n^4}\right),$$

we have

$$g(W) = g(Z') + \ln t + O\left(\frac{1}{m^3n^3}\right).$$

From (4.5.1) it follows that

$$g(Z) - g(W) = O\left(\frac{1}{m^3n^3}\right). \tag{4.5.2}$$

Next, we are going to exploit the strong concavity of g and use the following standard inequality:

if $g''(x) \leq -\alpha$ for some $\alpha > 0$ and all $a \leq x \leq b$ then

$$g\left(\frac{a+b}{2}\right) - \frac{1}{2}g(a) - \frac{1}{2}g(b) \ \geq \ \frac{\alpha(b-a)^2}{8}.$$

If for some i, j we have $|w_{ij} - z_{ij}| \geq (mn)^{-1} w_{ij}$, then in view of (4.5.2), for some point U on the interval connecting W and Z and all sufficiently large mn, we will have

$$g(U) > g(Z),$$

which is a contradiction. Thus

$$\left| \frac{z_i}{w_{ij}} - 1 \right| \ \leq \ \frac{1}{mn} \quad \text{for all} \quad i, j$$

and all sufficiently large mn. Since

$$\left| \frac{w_{ij}}{t z'_{ij}} - 1 \right| \ \leq \ \frac{3}{z'_{ij}} \ \leq \ \frac{3}{(mn)^4},$$

the proof follows. □

(4.6) Proof of Theorem 1.5. Without loss of generality we assume that $N \geq (mn)^7$ since the case of a polynomially bounded N is handled in Proposition 3.5.

Let us choose

$$t = \left\lfloor \frac{N}{(mn)^6} \right\rfloor$$

and consider the t-scaling map $\mathcal{T} : \Sigma(R, C) \longrightarrow \Sigma(R', C')$. Since margins (R, C) are δ-smooth, we have

$$(mn)^6 \ \leq \ N' \ \leq \ (mn)^7 \quad \text{and} \quad r'_i, c'_j \ \geq \ (mn)^4 \quad \text{for all} \quad i, j$$

and all sufficiently large $n \geq m$.

Let us choose $0 < \delta_1 < \delta$. It follows by (4.2.2) that the margins (R', C') are δ_1-smooth for all sufficiently large $n \geq m$. Let Z' be the typical table of (R', C'), $Z' = \left(z'_{ij}\right)$. By Corollary 2.5,

$$z'_{ij} \ \geq \ (\delta_1)^3 \frac{N'}{mn}.$$

Therefore, for all sufficiently large $m + n$ we have

$$z'_{ij} \ \geq \ (mn)^4 + 3.$$

The result now follows by Lemmas 4.4, 4.5, and Proposition 3.5 applied to (R', C').
□

Acknowledgment

The author is grateful to Alexander Yong for asking in what sense the "typical table" introduced in [Ba09] and named so in [B+08] was typical. This paper is an attempt to answer that question. After the first version of this paper was written, John Hartigan pointed out to connections with the maximum entropy principle and suggested Theorem 1.7 to the author (cf. [BH09]), which led to a substantial simplification of the original proof and some strengthening of the main result, Theorem 1.5.

References

[Ba02] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics, 54, American Mathematical Society, Providence, RI, 2002.

[Ba08] A. Barvinok, *On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries*, preprint arXiv:0806.1480 (2008).

[B+08] A. Barvinok, Z. Luria, A. Samorodnitsky, and A. Yong, *An approximation algorithm for counting contingency tables*, preprint `arXiv:0803.3948`, Random Structures & Algorithms, to appear (2008).

[Ba09] A. Barvinok, *Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes*, International Mathematics Research Notices **2009** (2009), 348–385.

[BH09] A. Barvinok and J.A. Hartigan, *Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes*, preprint `arXiv:0903.5223` (2009).

[C+06] M. Cryan, M. Dyer, L.A. Goldberg, M. Jerrum, and R. Martin, *Rapidly mixing Markov chains for sampling contingency tables with a constant number of rows*, SIAM Journal on Computing **36** (2006), 247–278.

[DE85] P. Diaconis and B. Efron, *Testing for independence in a two-way table: new interpretations of the chi-square statistic. With discussions and with a reply by the authors*, The Annals of Statistics **13** (1985), 845–913.

[DG95] P. Diaconis and A. Gangolli, *Rectangular arrays with fixed margins*, Discrete probability and algorithms (Minneapolis, MN, 1993), The IMA Volumes in Mathematics and its Applications, 72, Springer, New York, 1995, pp. 15–41.

[D+97] M. Dyer, R. Kannan, and J. Mount, *Sampling contingency tables*, Random Structures & Algorithms **10** (1997), 487–506.

[Go63] I. J. Good, *Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables*, The Annals of Mathematical Statistics **34** (1963), 911-934.

[GM08] C. Greenhill and B.D. McKay, *Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums*, Advances in Applied Mathematics **41** (2008), 459–481.

[Le01] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, 89, American Mathematical Society, Providence, RI, 2001.

[Ne69] P.E. O'Neil, *Asymptotics and random matrices with row-sum and column-sum restrictions*, Bulletin of the American Mathematical Society **75** (1969), 1276–1282.

[NN94] Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics, 13, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

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

E-mail address: `barvinok@umich.edu`