MATRICES WITH PRESCRIBED ROW AND COLUMN SUMS

Alexander Barvinok

October 2010

Abstract. This is a survey of the recent progress and open questions on the structure of the sets of 0-1 and non-negative integer matrices with prescribed row and column sums. We discuss cardinality estimates, the structure of a random matrix from the set, discrete versions of the Brunn-Minkowski inequality and the statistical dependence between row and column sums.

1. Introduction

Let $R=(r_1, \ldots, r_m)$ and $C=(c_1, \ldots, c_n)$ be positive integer vectors such that

$$r_1 + \ldots + r_m = c_1 + \ldots + c_n = N. \tag{1.1}$$

We consider the set $A_0(R, C)$ of all $m \times n$ matrices $D=(d_{ij})$ with 0-1 entries, row sums R and column sums C:

$$A_0(R,C)=\Bigg\{D=(d_{ij}): \quad \begin{aligned} &\sum_{j=1}^n d_{ij} = r_i \quad \text{for} \quad i=1, \ldots, m \\ &\sum_{i=1}^m d_{ij} = c_j \quad \text{for} \quad j=1, \ldots, n \\ &d_{ij} \in \{0, 1\} \end{aligned} \Bigg\}.$$

Key words and phrases. 0-1 matrix, integer matrix, random matrix, permanent, Brunn - Minkowski inequality, Central Limit Theorem.

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

Typeset by $\mathcal{AMS}$-TEX

We also consider the set $A_+(R,C)$ of non-negative integer $m \times n$ matrices with row sums R and column sums C:

$$A_+(R,C) = \Bigg\{ D = (d_{ij}) : \qquad \sum_{j=1}^{n} d_{ij} = r_i \quad \text{for} \quad i = 1, \dots, m$$
$$\sum_{i=1}^{m} d_{ij} = c_j \quad \text{for} \quad j = 1, \dots, n$$
$$d_{ij} \in \mathbb{Z}_+ \Bigg\}.$$

Vectors R and C are called *margins* of matrices from $A_0(R,C)$ and $A_+(R,C)$. We reserve notation N for the sums of the coordinates of R and C in (1.1) and write $|R| = |C| = N$.

While the set $A_+(R,C)$ is non-empty as long as the balance condition (1.1) is satisfied, a result of Gale and Ryser (see, for example, Section 6.2 of [BR91]) provides a necessary and sufficient criterion for set $A_0(R,C)$ to be non-empty. Let us assume that

$$m \ \geq \ c_1 \ \geq \ c_2 \ \geq \ \dots \ \geq \ c_n \ \geq \ 0 \quad \text{and that}$$
$$n \ \geq \ r_i \ \geq \ 0 \quad \text{for} \quad i = 1, \dots, n.$$

Set $A_0(R,C)$ is not empty if and only if (1.1) holds and

$$\sum_{i=1}^{m} \min\{r_i, k\} \ \geq \ \sum_{j=1}^{k} c_j \quad \text{for} \quad k = 1, \dots, n.$$

Assuming that $A_0(R,C) \neq \emptyset$, we are interested in the following questions:

• What is the cardinality $|A_0(R,C)|$ of $A_0(R,C)$ and the cardinality $|A_+(R,C)|$ of $A_+(R,C)$?

• Let us us consider $A_0(R,C)$ and $A_+(R,C)$ as finite probability spaces with the uniform measure. What a random matrix $D \in A_0(R,C)$ and a random matrix $D \in A_+(R,C)$ are likely to look like?

The paper is organized as follows.

In Section 2 we estimate of $|A_0(R,C)|$ within an $(mn)^{O(m+n)}$ factor and in Section 3 we estimate $|A_+(R,C)|$ within an $N^{O(m+n)}$ factor. In all but very sparse cases this way we obtain asymptotically exact estimates of $\ln|A_0(R,C)|$ and $\ln|A_+(R,C)|$ respectively. The estimate of Section 2 is based on a representation of $|A_0(R,C)|$ as the permanent of a certain $mn \times mn$ matrix of 0's and 1's, while the estimate of Section 3 is based on a representation of $|A_+(R,C)|$ as the expectation of the permanent of a certain $N \times N$ random matrix with exponentially distributed entries. In the proofs, the crucial role is played by the van der Waerden inequality for permanents of doubly stochastic matrices. The cardinality estimates are obtained as solutions to simple convex optimization problems and hence are efficiently computable, although they cannot be expressed by a "closed formula" in the margins (R, C). Our method is sufficiently robust as the same approach can be applied to estimate the cardinality of the set of matrices with prescribed margins *and* with 0's in prescribed positions.

In Sections 4 and 5 we discuss some consequences of the formulas obtained in Sections 2 and 3. In particular, in Section 4, we show that the numbers $|A_0(R, C)|$ and $|A_+(R, C)|$ are both approximately log-concave as functions of the margins (R, C). We note an open question whether these numbers are genuinely log-concave and give some, admittedly weak, evidence that it may be the case. In Section 5, we discuss statistical dependence between row and column sums. Namely, we consider finite probability spaces of $m \times n$ non-negative integer or 0-1 matrices with the total sum N of entries and two events in those spaces: event $\mathcal{R}$ consisting of the matrices with row sums R and event $\mathcal{C}$ consisting of the matrices with column sums C. It turns out that 0-1 and non-negative integer matrices exhibit opposite types of behavior. Assuming that the margins R and C are sufficiently far away from sparse and uniform, we show that for 0-1 matrices the events $\mathcal{R}$ and $\mathcal{C}$ repel each other (events $\mathcal{R}$ and $\mathcal{C}$ are negatively correlated) while for non-negative integer matrices they attract each other (the events are positively correlated).

In Section 6, we discuss what random matrices $D \in A_0(R, C)$ and $D \in A_+(R, C)$ look like. We show that in many respects, a random matrix $D \in A_0(R, C)$ behaves like an $m \times n$ matrix X of independent Bernoulli random variables such that $\mathbf{E}\, X = Z_0$ where Z_0 is a certain matrix, called the maximum entropy matrix, with row sums R, column sums C and entries between 0 and 1. It turns out that Z_0 is the solution to an optimization problem, which is convex dual to the optimization problem of Section 2 used to estimate $|A_0(R, C)|$. On the other hand, a random matrix $D \in A_+(R, C)$ in many respects behaves like an $m \times n$ matrix X of independent geometric random variables such that $\mathbf{E}\, X = Z_+$ where Z_+ is a certain matrix, also called the maximum entropy matrix, with row sums R, column sums C and non-negative entries. It turns out that Z_+ is the solution to an optimization problem which is convex dual to the optimization problem of Section 3 used to estimate $|A_+(R, C)|$. It follows that in various natural metrics matrices $D \in A_0(R, C)$ concentrate about Z_0 while matrices $D \in A_+(R, C)$ concentrate about Z_+. We note some open questions on whether individual entries of random $D \in A_0(R, C)$ and random $D \in A_+(R, C)$ are asymptotically Bernoulli, respectively geometric, with the expectations read off from Z_0 and Z_+.

In Section 7, we discuss asymptotically exact formulas for $|A_0(R, C)|$ and $|A_+(R, C)|$. Those formulas are established under essentially more restrictive conditions than cruder estimates of Sections 2 and 3. We assume that the entries of the maximum entropy matrices Z_0 and Z_+ are within a constant factor, fixed in advance, of each other. Recall that matrices Z_0 and Z_+ characterize the typical behavior of random matrices $D \in A_0(R,C)$ and $D \in A_+(R,C)$ respectively. In the case of 0-1 matrices our condition basically means that the margins (R,C) lie sufficiently deep inside the region defined by the Gale-Ryser inequalities. As the margins approach the boundary, the number $|A_0(R,C)|$ gets volatile and hence cannot be expressed by an analytic formula like the one described in Section 7. The situation with non-negative integer matrices is less clear. It is plausible that the number $|A_+(R,C)|$ experiences some volatility when some entries of Z_+ become abnormally large, but we don't have a proof of that happenning.

In Section 8, we mention some possible ramifications, such as enumeration of higher-order tensors and graphs with given degree sequences.

The paper is a survey and although we don't provide complete proofs, we often sketch main ideas of our approach.

2. The logarithmic asymptotic for the number of 0-1 matrices

The following result is proven in [Ba10a].

(2.1) Theorem. *Given positive integer vectors*

$$R=(r_1, \ldots, r_m) \quad \textit{and} \quad C=(c_1, \ldots, c_n),$$

let us define the function

$$F_0(\mathbf{x}, \mathbf{y}) = \left(\prod_{i=1}^m x_i^{-r_i}\right)\left(\prod_{j=1}^n y_j^{-c_j}\right)\left(\prod_{i,j}(1+x_iy_j)\right)$$
$$\textit{for} \quad \mathbf{x}=(x_1, \ldots, x_m) \quad \textit{and} \quad \mathbf{y}=(y_1, \ldots, y_n)$$

and let

$$\alpha_0(R,C) = \inf_{\substack{x_1, \ldots, x_m > 0 \\ y_1, \ldots, y_n > 0}} F_0(\mathbf{x}, \mathbf{y}).$$

Then the number $A_0(R,C)$ of $m \times n$ zero-one matrices with row sums R and column sums C satisfies

$$\alpha_0(R,C) \ \geq \ |A_0(R,C)| \ \geq \ \frac{(mn)!}{(mn)^{mn}}\left(\prod_{i=1}^m \frac{(n-r_i)^{n-r_i}}{(n-r_i)!}\right)\left(\prod_{j=1}^n \frac{c_j^{c_j}}{c_j!}\right)\alpha_0(R,C).$$

Using Stirling's formula,

$$\frac{s!}{s^s} = \sqrt{2\pi s}e^{-s}\left(1+O\left(\frac{1}{s}\right)\right),$$

one can notice that the ratio between the upper and lower bounds is $(mn)^{O(m+n)}$. Indeed, the "e^{-s}" terms cancel each other out, since

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

Thus, for sufficiently dense 0-1 matrices, where we have $|A_0(R,C)|=2^{\Omega(mn)}$, we have an asymptotically exact formula

$$\ln|A_0(R,C)| \;\approx\; \ln\alpha_0(R,C) \quad \text{as} \quad m,n\longrightarrow+\infty.$$

(2.2) A convex version of the optimization problem. Let us substitute

$$x_i=e^{s_i} \quad \text{for} \quad i=1,\ldots,m \quad \text{and} \quad y_j=e^{t_j} \quad \text{for} \quad j=1,\ldots,n$$

in $F_0(\mathbf{x},\mathbf{y})$. Denoting

$$\begin{gathered} G_0(\mathbf{s},\mathbf{t})=-\sum_{i=1}^{m} r_i s_i-\sum_{j=1}^{n} t_j c_j+\sum_{i,j}\ln\left(1+e^{s_i+t_j}\right) \\ \text{for} \quad \mathbf{s}=(s_1,\ldots,s_m) \quad \text{and} \quad \mathbf{t}=(t_1,\ldots,t_n), \end{gathered} \tag{2.2.1}$$

we obtain

$$\ln\alpha_0(R,C)=\inf_{\substack{s_1,\ldots,s_m\\ t_1,\ldots,t_n}} G_0(\mathbf{s},\mathbf{t}).$$

We observe that $G_0(\mathbf{s},\mathbf{t})$ is a convex function on $\mathbb{R}^{m+n}$. In particular, one can compute the infimum of G_0 efficiently by using interior point methods, see, for example, [NN94].

(2.3) Sketch of proof of Theorem 2.1. The upper bound for $|A_0(R,C)|$ is immediate: it follows from the expansion

$$\prod_{ij}(1+x_i y_j)=\sum_{R,C}|A_0(R,C)|\mathbf{x}^R\mathbf{y}^C,$$

where

$$\mathbf{x}^R=x_1^{r_1}\cdots x_m^{r_m} \quad \text{and} \quad \mathbf{y}^C=y_1^{c_1}\cdots y_n^{c_n}$$

for $R=(r_1,\ldots,r_m)$ and $C=(c_1,\ldots,c_n)$ and the sum is taken over all pairs of non-negative integer vectors $R=(r_1,\ldots,r_m)$ and $C=(c_1,\ldots,c_n)$ such that $r_1+\ldots+r_m=c_1+\ldots+c_n\leq mn$.

To prove the lower bound, we express $|A_0(R,C)|$ as the permanent of an $mn\times mn$ matrix. Recall that the *permanent* of a $k\times k$ matrix $B=(b_{ij})$ is defined by

$$\operatorname{per}B=\sum_{\sigma\in S_k}\prod_{i=1}^{k} b_{i\sigma(i)},$$

where the sum is taken over the symmetric group S_k of all permutations σ of the set $\{1, \ldots, k\}$, see, for example, Chapter 11 of [LW01]. One can show, see [Bar10a] for details, that

$$(2.3.1) \qquad |A_0(R,C)| = \left(\prod_{i=1}^{m} \frac{1}{(n-r_i)!}\right)\left(\prod_{j=1}^{n} \frac{1}{c_j!}\right) \operatorname{per} B,$$

where B is the $mn \times mn$ matrix of the following structure:

the rows of B are split into distinct $m+n$ blocks, the m blocks of type I having $n-r_1, \ldots, n-r_m$ rows respectively and n blocks of type II having $c_1, \ldots, c_n$ rows respectively;

the columns of B are split into m distinct blocks of n columns each;

for $i=1, \ldots, m$, the entry of B that lies in a row from the i-th block of rows of type I and a column from the i-th block of columns is equal to 1;

for $i=1, \ldots, m$ and $j=1, \ldots, n$, the entry of B 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 1;

all other entries of B are 0.

Suppose that the infimum of function $G_0(\mathbf{s}, \mathbf{t})$ defined by (2.2.1) is attained at a particular point $\mathbf{s}=(s_1, \ldots, s_m)$ and $\mathbf{t}=(t_1, \ldots, t_n)$ (the case when the infimum is not attained is handled by an approximation argument). Let $x_i = \exp\{s_i\}$ for $i=1, \ldots, m$ and $y_j = \exp\{t_j\}$ for $j=1, \ldots, n$.

Setting the gradient of $G_0(\mathbf{s}, \mathbf{t})$ to 0, we obtain

$$(2.3.2) \qquad \begin{aligned} \sum_{j=1}^{n} \frac{x_i y_j}{1+x_i y_j} &= r_i \quad \text{for} \quad i=1, \ldots, m \\ \sum_{i=1}^{m} \frac{x_i y_j}{1+x_i y_j} &= c_j \quad \text{for} \quad j=1, \ldots, n. \end{aligned}$$

Let us consider a matrix B' obtained from matrix B as follows:

for $i=1, \ldots, m$ we multiply every row of B in the i-th block of type I by

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

for $j=1, \ldots, n$, we multiply every row of B in the j-th block of type II by

$$\frac{y_j}{c_j};$$

for $i=1, \ldots, m$ and $j=1, \ldots, n$ we multiply the j-th column in the i-th block of columns of B by

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

Then

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

On the other hand, equations (2.3.2) imply that the row and column sums of B' are equal to 1, that is, B' is *doubly stochastic*. Applying the van der Waerden bound for permanents of doubly stochastic matrices, see, for example, Chapter 12 of [LW01], we conclude that

$$\operatorname{per} B' \ \geq \ \frac{(mn)!}{(mn)^{mn}},$$

which, together with (2.3.1) completes the proof. □

One can prove a version of Theorem 2.1 for 0-1 matrices with prescribed row and column sums *and* prescribed zeros in some positions.

3. The logarithmic asymptotics for the number of non-negative integer matrices

The following result is proven in [Ba09].

(3.1) 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*

$$F_+(\mathbf{x}, \mathbf{y}) = \left(\prod_{i=1}^m x_i^{-r_i}\right) \left(\prod_{j=1}^n y_j^{-c_j}\right) \left(\prod_{ij} \frac{1}{1-x_i y_j}\right)$$

$$\text{for} \quad \mathbf{x}=(x_1, \ldots, x_m) \quad \text{and} \quad \mathbf{y}=(y_1, \ldots, y_n).$$

Then $F_+(\mathbf{x}, \mathbf{y})$ attains its minimum

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

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

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

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

For sufficiently dense matrices, where

$$\min_{i=1,\ldots,m} r_i = \Omega(n) \quad \text{and} \quad \min_{j=1,\ldots,n} c_j = \Omega(m)$$

we have $|A_+(R,C)| = (N/mn)^{\Omega(mn)}$ and hence we obtain an asymptotically exact formula

$$\ln |A_+(R,C)| \approx \ln \alpha_+(R,C) \quad \text{as} \quad m,n \longrightarrow +\infty.$$

(3.2) A convex version of the optimization problem. Let us substitute

$$x_i = e^{-s_i} \quad \text{for} \quad i=1,\ldots,m \quad \text{and} \quad y_j = e^{-t_j} \quad \text{for} \quad j=1,\ldots,n$$

in $F_+(\mathbf{x},\mathbf{y})$. Denoting

$$G_+(\mathbf{s},\mathbf{t}) = \sum_{i=1}^m r_i s_i + \sum_{j=1}^n t_j c_j - \sum_{i,j} \ln\left(1 - e^{-s_i - t_j}\right) \tag{3.2.1}$$
$$\text{for} \quad \mathbf{s} = (s_1,\ldots,s_m) \quad \text{and} \quad \mathbf{t} = (t_1,\ldots,t_n),$$

we obtain

$$\ln \alpha_+(R,C) = \min_{\substack{s_1,\ldots,s_m>0 \\ t_1,\ldots t_n>0}} G_+(\mathbf{s},\mathbf{t}).$$

We observe that $G_+(\mathbf{s},\mathbf{t})$ is a convex function on $\mathbb{R}^{m+n}$. In particular, one can compute the minimum of G_+ efficiently by using interior point methods [NN94].

(3.3) Sketch of proof of Theorem 3.1. The upper bound for $|A_+(R,C)|$ follows immediately from the expansion

$$\prod_{ij} \frac{1}{1 - x_i y_j} = \sum_{R,C} |A_+(R,C)| \mathbf{x}^R \mathbf{y}^C \quad \text{for} \quad 0 < x_1,\ldots,x_m,y_1,\ldots,y_n < 1$$

where

$$\mathbf{x}^R = x_1^{r_1} \cdots x_m^{r_m} \quad \text{and} \quad \mathbf{y}^C = y_1^{c_1} \cdots y_n^{c_n}$$

for $R = (r_1,\ldots,r_m)$ and $C = (c_1,\ldots,c_n)$ and the sum is taken over all pairs of non-negative integer vectors $R = (r_1,\ldots,r_m)$ and $C = (c_1,\ldots,c_n)$ such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n$.

To prove the lower bound, we express $|A_+(R,C)|$ as the integral of the permanent of an $N \times N$ matrix with variable entries. For an $m \times n$ matrix $Z = (z_{ij})$ we define the $N \times N$ matrix $B(Z)$ as follows:

the rows of $B(Z)$ are split into m distinct blocks of sizes $r_1,\ldots,r_m$ respectively;

the columns of $B(Z)$ are split into n distinct blocks of sizes $c_1,\ldots,c_n$ respectively;

for $i=1,\ldots,m$ and $j=1,\ldots,n$, the entry of $B(Z)$ that lies in a row from the i-th block of rows and in a column from the j-th block of columns is z_{ij}.

Then there is a combinatorial identity

$$\operatorname{per} B(Z) = \left(\prod_{i=1}^m r_i!\right)\left(\prod_{j=1}^n c_j!\right) \sum_{\substack{D\in A_+(R,C)\\ D=(d_{ij})}} \prod_{ij} \frac{z_{ij}^{d_{ij}}}{d_{ij}!},$$

cf. [Be74], which implies that

$$|A_+(R,C)| = \left(\prod_{i=1}^m \frac{1}{r_i!}\right)\left(\prod_{j=1}^n \frac{1}{c_j!}\right)\int_{\mathbb{R}_+^{mn}} \operatorname{per} B(Z) \exp\left\{-\sum_{ij} z_{ij}\right\}\ dZ.$$

Here the integral is taken over the set $\mathbb{R}_+^{mn}$ of $m\times n$ matrices Z with positive entries. Let $\Delta_{mn-1} \subset \mathbb{R}_+^{mn}$ be the standard $(mn-1)$-dimensional simplex defined by the equation

$$\sum_{ij} z_{ij} = 1.$$

Since $\operatorname{per} B(Z)$ is a homogeneous polynomial in Z of degree N, we have

$$(3.3.1)\quad |A_+(R,C)| = \frac{(N+mn-1)!}{\sqrt{mn}}\left(\prod_{i=1}^m \frac{1}{r_i!}\right)\left(\prod_{j=1}^n \frac{1}{c_j!}\right)\int_{\Delta_{mn-1}} \operatorname{per} B(Z)\ dZ,$$

where dZ is the Lebesgue measure on Δ_{mn-1} induced from $\mathbb{R}^{mn}$.

Let $\mathbf{s}=(s_1,\ldots,s_m)$ and $\mathbf{t}=(t_1,\ldots,t_n)$ be the minimum point of function $G_+(\mathbf{s},\mathbf{t})$ defined by (3.2.1). Let $x_i=\exp\{s_i\}$ for $i=1,\ldots,m$ and $y_j=\exp\{t_j\}$ for $j=1,\ldots,m$. Setting the gradient of $G_+(\mathbf{s},\mathbf{t})$ to 0, we obtain

$$(3.3.2)\qquad \begin{aligned} \sum_{j=1}^n \frac{x_i y_j}{1-x_i y_j} &= r_i \quad \text{for} \quad i=1,\ldots,m \\ \sum_{i=1}^m \frac{x_i y_j}{1-x_i y_j} &= c_j \quad \text{for} \quad j=1,\ldots,n. \end{aligned}$$

Let us consider the affine subspace $L\subset\mathbb{R}^{mn}$ of $m\times n$ matrices $Z=(z_{ij})$ defined by the system of equations

$$(3.3.3)\qquad \begin{aligned} \sum_{j=1}^n x_i y_j z_{ij} &= \frac{r_i}{N+mn} \quad \text{for} \quad i=1,\ldots,m \quad \text{and} \\ \sum_{i=1}^m x_i y_j z_{ij} &= \frac{c_j}{N+mn} \quad \text{for} \quad j=1,\ldots,n. \end{aligned}$$

We note that $\dim L = (m-1)(n-1)$.

Suppose that $Z \in \Delta_{mn-1} \cap L$ and consider the corresponding matrix $B(Z)$. If we multiply every row in the i-th block of rows by $x_i\sqrt{N+mn}/r_i$ and every column in the j-th block of columns by $y_j\sqrt{N+mn}/c_j$, by (3.3.3) we obtain a doubly stochastic matrix $B'(Z)$ for which we have $\operatorname{per} B'(Z) \geq N!/N^N$ by the van der Waerden inequality. Summarizing,

$$
\begin{aligned}
\operatorname{per} B(Z) \ &\geq \ \frac{N!}{N^N (N+mn)^N} \left(\prod_{i=1}^{m} r_i^{r_i}\right) \left(\prod_{j=1}^{n} c_j^{c_j}\right) \\
&\times \left(\prod_{i=1}^{m} x_i^{-r_i}\right) \left(\prod_{j=1}^{n} y_j^{-c_j}\right) \qquad \text{for all} \quad Z \in \Delta_{mn-1} \cap L.
\end{aligned}
\tag{3.3.4}
$$

It remains to show that the intersection $\Delta_{mn-1} \cap L$ is sufficiently large, so that the contribution of a neighborhood of the intersection to the integral (3.3.1) is sufficiently large. It follows by (3.3.2)–(3.3.3) that $\Delta_{mn-1} \cap L$ contains matrix $Z = (z_{ij})$ where

$$
z_{ij} = \frac{1}{(N+mn)\,(1-x_i y_j)} \quad \text{for all} \quad i, j.
$$

In [Ba09], we prove a geometric lemma which states that if $\Delta_{d-1} \subset \mathbb{R}^d_+$ is the standard $(d-1)$-dimensional simplex that is the intersection of the affine hyperplane H defined by the equation $x_1 + \ldots + x_d = 1$ and the positive orthant $x_1 > 0, \ldots, x_d > 0$ and if $L \subset H$ is an affine subspace of codimension k in H such that L contains a point $a \in \Delta_{d-1}$, $a = (\alpha_1, \ldots, \alpha_d)$, then for the volume of the intersection $\Delta_{d-1} \cap L$ we have the lower bound

$$
\operatorname{vol}_{d-k-1}\left(\Delta_{d-1} \cap L\right) \ \geq \ \frac{\gamma}{d!\,\omega_k} d^{d-\frac{1}{2}} \alpha_1 \ldots \alpha_d,
$$

where

$$
\omega_k = \frac{\pi^{k/2}}{\Gamma(k/2+1)}
$$

is the volume of the k-dimensional unit ball and $\gamma > 0$ is an absolute constant. Applying this estimate in our situation, we conclude that

$$
\operatorname{vol}_{mn-k}\left(\Delta_{mn-1} \cap L\right) \ \geq \ \frac{1}{(mn)^{O(m+n)}} \frac{e^{mn}}{(N+mn)^{mn}} \prod_{ij} \frac{1}{1-x_i y_j},
$$

where $k = m+n-1$ or $k = m+n$ depending whether or not L lies in the affine hyperplane $\sum_{ij} z_{ij} = 1$. This allows us to obtain a similar bound for the volume of a small neighborhood of the intersection $\Delta_{mn-1} \cap L$. Because $\operatorname{per} B(Z)$ is a homogeneous polynomial in Z of degree N, inequality (3.3.4) holds in the ϵ-neighborhood of the intersection $\Delta_{mn-1} \cap L$ for $\epsilon = N^{-O(m+n)}$ up to an $N^{O(m+n)}$ factor. Using it together with (3.3.1), we complete the proof of Theorem 3.1. □

One can prove a version of Theorem 3.1 for non-negative integer matrices with prescribed row and column sums *and* with prescribed zeros in some positions.

4. Discrete Brunn - Minkowski inequalities

Theorems 2.1 and 3.1 allow us to establish approximate log-concavity of the numbers $A_0(R,C)$ and $A_+(R,C)$.

For a non-negative integer vector $B=(b_1,\ldots,b_p)$, we denote

$$|B|=\sum_{i=1}^p b_i.$$

(4.1) Theorem. *Let $R_1,\ldots,R_p$ be positive integer m-vectors and let $C_1,\ldots,C_p$ be positive integer n-vectors such that $|R_1|=|C_1|,\ldots,|R_p|=|C_p|$.*

Let $\beta_1,\ldots,\beta_p\geq 0$ be real numbers such that $\beta_1+\ldots+\beta_p=1$ and such that $R=\beta_1R_1+\ldots+\beta_pR_p$ is a positive integer m-vector and $C=\beta_1C_1+\ldots+\beta_pC_p$ is a positive integer n-vector. Let $N=|R|=|C|$.

Then for some absolute constant $\gamma>0$ we have

(1)

$$(mn)^{\gamma(m+n)}\left|A_0(R,C)\right| \ \geq \ \prod_{k=1}^p \left|A_0\left(R_k,C_k\right)\right|^{\beta_k}$$

and

(2)

$$N^{\gamma(m+n)}\left|A_+(R,C)\right| \ \geq \ \prod_{k=1}^p \left|A_+\left(R_k,C_k\right)\right|^{\beta_k}.$$

Proof. Let us denote function F_0 of Theorem 2.1 for the pair (R_k,C_k) by F_k and for the pair (R,C) just by F. Then

$$F(\mathbf{x},\mathbf{y})=\prod_{k=1}^p F_k^{\beta_k}(\mathbf{x},\mathbf{y}) \tag{4.1.1}$$

and hence

$$\alpha_0(R,C) \ \geq \ \prod_{k=1}^p \left(\alpha_0\left(R_k,C_k\right)\right)^{\beta_k}.$$

Part (1) now follows by Theorem 2.1.

Similarly, we obtain (4.1.1) if we denote function F_+ of Theorem 3.1 for the pair (R_k,C_k) by F_k and for the pair (R,C) just by F. Hence

$$\alpha_+(R,C) \ \geq \ \prod_{k=1}^p \left(\alpha_+\left(R_k,C_k\right)\right)^{\beta_k}.$$

Part (2) now follows by Theorem 3.1. □

Teorem 2.1 implies a more precise estimate

$$\frac{(mn)^{mn}}{(mn)!}\left(\prod_{i=1}^{m}\frac{(n-r_i)!}{(n-r_i)^{n-r_i}}\right)\left(\prod_{j=1}^{n}\frac{c_j!}{c_j^{c_j}}\right)|A_0(R,C)| \ \geq \ \prod_{k=1}^{p}|A_0\left(R_k,C_k\right)|^{\beta_k},$$

where $R=(r_1,\ldots,r_m)$ and $C=(c_1,\ldots,c_n)$.

In [Ba07] a more precise estimate

$$\frac{N^N}{N!}\min\left\{\prod_{i=1}^{m}\frac{r_i!}{r_i^{r_i}}, \quad \prod_{j=1}^{n}\frac{c_j!}{c_j^{c_j}}\right\}|A_+(R,C)| \ \geq \ \prod_{k=1}^{p}|A_+\left(R_k,C_k\right)|^{\beta_k}$$

is proven under the additional assumption that $|R_k|=|C_k|=N$ for $k=1,\ldots,p$.

Theorem 4.1 raises a natural question whether stronger inequalities hold.

(4.2) Brunn-Minkowski inequalities.

(4.2.1) Question. Is it true that under the conditions of Theorem 4.1 we have

$$|A_0(R,C)| \ \geq \ \prod_{k=1}^{p}|A_0\left(R_k,C_k\right)|^{\beta_k}?$$

(4.2.2) Question. Is it true that under the conditions of Theorem 4.1 we have

$$|A_+(R,C)| \ \geq \ \prod_{k=1}^{p}|A_+\left(R_k,C_k\right)|^{\beta_k}?$$

Should they hold, inequalities of (4.2.1) and (4.2.2) would be natural examples of discrete Brunn-Minkowski inequalities, see [Ga02] for a survey.

Some known simpler inequalities are consistent with the inequalities of (4.2.1)–(4.2.2). Let $X=(x_1,\ldots,x_p)$ and $Y=(y_1,\ldots,y_p)$ be non-negative integer vectors such that

$$x_1 \geq x_2 \geq \ldots \geq x_p \quad \text{and} \quad y_1 \geq y_2 \geq \ldots \geq y_p.$$

We say that X *dominates* Y if

$$\sum_{i=1}^{k} x_i \ \geq \ \sum_{i=1}^{k} y_i \quad \text{for} \quad k=1,\ldots,p-1 \quad \text{and} \quad \sum_{i=1}^{p} x_i = \sum_{i=1}^{p} y_i.$$

Equivalently, X dominates Y if Y is a convex combination of vectors obtained from X by permutations of coordinates.

One can show that

$$(4.2.3) \qquad |A_0(R,C)| \ \geq \ |A_0\left(R',C'\right)| \quad \text{and} \quad |A_+(R,C)| \ \geq \ |A_+\left(R',C'\right)|$$

provided R' dominates R and C' dominates C, see Chapter 16 of [LW01] and [Ba07]. Inequalities (4.2.3) are consistent with the inequalities of (4.2.1) and (4.2.2).

5. Dependence between row and column sums

The following attractive "independence heuristic" for estimating $|A_0(R,C)|$ and $|A_+(R,C)|$ was discussed by Good [Go77] and by Good and Crook [GC76].

(5.1) The independence heuristic. Let us consider the set of all $m \times n$ matrices $D = (d_{ij})$ with 0-1 entries and the total sum N of entries as a finite probability space with the uniform measure. Let us consider the event $\mathcal{R}_0$ consisting of the matrices with the row sums $R = (r_1, \ldots, r_m)$ and the event $\mathcal{C}_0$ consisting of the matrices with the column sums $C = (c_1, \ldots, c_n)$. Then

$$\mathbf{Pr}\ (\mathcal{R}_0) = \binom{mn}{N}^{-1} \prod_{i=1}^{m} \binom{n}{r_i} \quad \text{and} \quad \mathbf{Pr}\ (\mathcal{C}_0) = \binom{mn}{N}^{-1} \prod_{j=1}^{n} \binom{m}{c_j}.$$

In addition,

$$A_0(R,C) = \mathcal{R}_0 \cap \mathcal{C}_0.$$

If we assume that events $\mathcal{R}_0$ and $\mathcal{C}_0$ are independent, we obtain the following *independence estimate*

$$I_0(R,C) = \binom{mn}{N}^{-1} \prod_{i=1}^{m} \binom{n}{r_i} \prod_{j=1}^{n} \binom{m}{c_j} \tag{5.1.1}$$

for the number $|A_0(R,C)|$ of 0-1 matrices with row sums R and column sums C.

Similarly, let us consider the set of all $m \times n$ matrices $D = (d_{ij})$ with non-negative integer entries and the total sum N of entries as a finite probability space with the uniform measure. Let us consider the event $\mathcal{R}_+$ consisting of the matrices with the row sums $R = (r_1, \ldots, r_m)$ and the event $\mathcal{C}_+$ consisting of the matrices with the column sums $C = (c_1, \ldots, c_n)$. Then

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

We have

$$A_+(R,C) = \mathcal{R}_+ \cap \mathcal{C}_+.$$

If we assume that events $\mathcal{R}_+$ and $\mathcal{C}_+$ are independent, we obtain the *independence estimate*

$$I_+(R,C) = \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{r_i+n-1}{n-1} \prod_{j=1}^{n} \binom{c_j+m-1}{m-1}. \tag{5.1.2}$$

Interestingly, the independence estimates $I_0(R, C)$ and $I_+(R, C)$ provide reasonable approximations to $|A_0(R, C)|$ and $|A_+(R, C)|$ respectively in the following two cases:

in the case of equal margins, when

$$r_1 = \ldots = r_m = r \quad \text{and} \quad c_1 = \ldots = c_n = c,$$

see [C+08] and [C+07]

in the sparse case, when

$$\max_{i=1,\ldots,m} r_i \ll n \quad \text{and} \quad \max_{j=1,\ldots,n} c_j \ll m,$$

see [G+06] and [GM08].

We will see in Section 5.4 that the independence estimates provide the correct logarithmic asymptotics in the case when all row sums are equal *or* all column sums are equal. However, if both row and column sums are sufficiently far away from being uniform and sparse, the independence estimates, generally speaking, provide poor approximations. Moreover, in the case of 0-1 matrices the independence estimate $I_0(R, C)$ typically grossly overestimates $|A_0(R, C)|$ while in the case of non-negative integer matrices the independence estimate $I_+(R, C)$ typically grossly underestimates $|A_+(R, C)|$. In other words, for typical margins R and C the events $\mathcal{R}_0$ and $\mathcal{C}_0$ *repel* each other (the events are negatively correlated) while events $\mathcal{R}_+$ and $\mathcal{C}_+$ *attract* each other (the events are positively correlated). To see why this is the case, we write the estimates $\alpha_0(R, C)$ of Theorem 2.1 and $\alpha_+(R, C)$ of Theorem 3.1 in terms of entropy.

The following result is proven in [Ba10a].

(5.2) Lemma. *Let $P_0(R, C)$ be the polytope of all $m \times n$ matrices $X = (x_{ij})$ with row sums R, column sums C and such that $0 \leq x_{ij} \leq 1$ for all i and j. Suppose that polytope $P_0(R, C)$ has a non-empty interior, that is contains a matrix $Y = (y_{ij})$ such that $0 < y_{ij} < 1$ for all i and j. Let us define a function $h : P_0(R, C) \longrightarrow \mathbb{R}$ by*

$$h(X) = \sum_{i,j} x_{ij} \ln \frac{1}{x_{i,j}} + (1 - x_{ij}) \ln \frac{1}{1 - x_{ij}} \quad \textit{for} \quad X \in P_0(R, C).$$

Then h is a strictly concave function on of $P_0(R, C)$ and hence attains its maximum on $P(R, C)$ at a unique matrix $Z_0 = (z_{ij})$, which we call the maximum entropy matrix. Moreover,

1. *We have $0 < z_{ij} < 1$ for all i and j;*
2. *The infimum $\alpha_0(R, C)$ of Theorem 2.1 is attained at some particular point $(\mathbf{x}, \mathbf{y})$;*
3. *We have $\alpha_0(R, C) = e^{h(Z_0)}$.*

Sketch of Proof. It is straightforward to check that h is strictly concave and that

$$\frac{\partial}{\partial x_{ij}} h(X) = \ln \frac{1 - x_{ij}}{x_{ij}}.$$

In particular, the (right) derivative at $x_{ij} = 0$ is $+\infty$, the (left) derivative at $x_{ij} = 1$ is $-\infty$ and the derivative for $0 < x_{ij} < 1$ is finite. Hence the maximum entropy matrix Z_0 must have all entries strictly between 0 and 1, since otherwise we can increase the value of h by perturbing Z_0 in the direction of a matrix Y from the interior of $P_0(R, C)$. This proves Part (1).

The Lagrange optimality conditions imply that

$$\ln \frac{1 - z_{ij}}{z_{ij}} = -\lambda_i - \mu_j \quad \text{for all} \quad i, j$$

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

$$z_{ij} = \frac{e^{\lambda_i + \mu_j}}{1 + e^{\lambda_i + \mu_j}} \quad \text{for all} \quad i, j. \tag{5.2.1}$$

In particular,

$$\begin{aligned} &\sum_{i=1}^{m} \frac{e^{\lambda_i + \mu_j}}{1 + e^{\lambda_i + \mu_j}} = c_j \quad \text{for} \quad j = 1, \ldots, n \quad \text{and} \\ &\sum_{j=1}^{n} \frac{e^{\lambda_i + \mu_j}}{1 + e^{\lambda_i + \mu_j}} = r_i \quad \text{for} \quad i = 1, \ldots, m. \end{aligned} \tag{5.2.2}$$

Equations (5.2.2) imply that point $\mathbf{s} = (\lambda_1, \ldots, \lambda_m)$ and $\mathbf{t} = (\mu_1, \ldots, \mu_n)$ is a critical point of function $G_0(\mathbf{s}, \mathbf{t})$ defined by (2.2.1) and hence the infimum $\alpha_0(R, C)$ of $F_0(\mathbf{x}, \mathbf{y})$ is attained at $x_i = e^{\lambda_i}$ for $i = 1, \ldots, m$ and $y_j = e^{\mu_j}$ for $j = 1, \ldots, n$. Hence Part (2) follows. Using (5.2.1) it is then straightforward to check that $F_0(\mathbf{x}, \mathbf{y}) = e^{h(Z_0)}$ for the minimum point $(\mathbf{x}, \mathbf{y})$. $\square$

We note that

$$h(x) = x \ln \frac{1}{x} + (1 - x) \ln \frac{1}{x} \quad \text{for} \quad 0 \leq x \leq 1$$

is the entropy of the Bernoulli random variable with expectation x, see Section 6.

The following result is proven in [Ba09].

(5.3) Lemma. *Let $P_+(R, C)$ be the polytope of all non-negative $m \times n$ matrices $X = (x_{ij})$ with row sums R and column sums C. Let us define a function $g : P_+(R, C) \longrightarrow \mathbb{R}$ by*

$$g(X) = \sum_{i,j} (x_{ij} + 1) \ln (1 + x_{ij}) - x_{ij} \ln x_{ij} \quad \textit{for} \quad X \in P_+(R, C).$$

Then g is a strictly concave function on $P_+(R,C)$ and hence attains its maximum on $P_+(R,C)$ at a unique matrix $Z_+ = (z_{ij})$, which we call the maximum entropy matrix. Moreover,

(1) *We have $z_{ij} > 0$ for all i,j and*
(2) *For the minimum $\alpha_+(R,C)$ of Theorem 3.1, we have $\alpha_+(R,C) = e^{g(Z_+)}$.*

Sketch of Proof. It is straightforward to check that g is strictly concave and that

$$\frac{\partial}{\partial x_{ij}} g(X) = \ln \frac{1+x_{ij}}{x_{ij}} \quad \text{for all} \quad i,j.$$

In particular, the (left) derivative is $+\infty$ for $x_{ij} = 0$ and finite for every $x_{ij} > 0$. Since $P_+(R,C)$ contains an interior point (for example, matrix $Y = (y_{ij})$ with $y_{ij} = r_i c_j / N$), arguing as in the proof of Lemma 5.2, we obtain Part (1).

The Lagrange optimality conditions imply that

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

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

$$z_{ij} = \frac{e^{-\lambda_i - \mu_j}}{1 - e^{-\lambda_i - \mu_j}} \quad \text{for all} \quad i,j. \tag{5.3.1}$$

In particular,

$$\begin{gathered}\sum_{i=m}^{n} \frac{e^{-\lambda_i - \mu_j}}{1 + e^{-\lambda_i - \mu_j}} = c_j \quad \text{for} \quad j = 1, \ldots, n \quad \text{and} \\ \sum_{j=1}^{n} \frac{e^{-\lambda_i - \mu_j}}{1 + e^{-\lambda_i - \mu_j}} = r_i \quad \text{for} \quad i = 1, \ldots, m.\end{gathered} \tag{5.3.2}$$

Equations (5.3.2) imply that the point $\mathbf{s} = (\lambda_1, \ldots, \lambda_m)$ and $\mathbf{t} = (\mu_1, \ldots, \mu_n)$ is a critical point of function $G_+(\mathbf{s}, \mathbf{t})$ defined by (3.2.1) and hence the minimum $\alpha_+(R,C)$ of $F_+(\mathbf{x}, \mathbf{y})$ is attained at $x_i = e^{\lambda_i}$ for $i = 1, \ldots, m$ and $y_j = e^{\mu_j}$ for $j = 1, \ldots, n$. Using (5.3.1), it is then straightforward to check that $F_+(\mathbf{x}, \mathbf{y}) = e^{h(Z_+)}$ for the minimum point $(\mathbf{x}, \mathbf{y})$. □

We note that

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

is the entropy of the geometric random variable with expectation x, see Section 6.

(5.4) Estimates of the cardinality via entropy. Let

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

be the entropy function defined on k-tuples (probability distributions) $p_1, \ldots, p_k$ such that $p_1 + \ldots + p_k = 1$ and $p_i \geq 0$ for $i = 1, \ldots, k$. Assuming that polytope $P_0(R, C)$ of Lemma 5.2 has a non-empty interior, we can write

$$\begin{aligned}\ln \alpha_0(R,C) =& N\mathbf{H}\left(\frac{z_{ij}}{N};\ i,j\right) + (mn-N)\mathbf{H}\left(\frac{1-z_{ij}}{mn-N};\ i,j\right)\\ &- N\ln N - (mn-N)\ln(mn-N),\end{aligned}$$

where $Z_0 = (z_{ij})$ is the maximum entropy matrix. On the other hand, for the independence estimate (5.1.1), we have

$$\begin{aligned}\ln I_0(R,C) =& N\mathbf{H}\left(\frac{r_i}{N};\ i\right) + (mn-N)\mathbf{H}\left(\frac{n-r_i}{mn-N};\ i\right)\\ &+ N\mathbf{H}\left(\frac{c_j}{N};\ j\right) + (mn-N)\mathbf{H}\left(\frac{m-c_j}{mn-N};\ j\right)\\ &- N\ln N - (mn-N)\ln(mn-N) + O\big((m+n)\ln(mn)\big).\end{aligned}$$

Using the inequality which relates the entropy of a distribution and the entropy of its margins, see, for example, [Kh57], we obtain

$$\mathbf{H}\left(\frac{z_{ij}}{N};\ i,j\right) \ \leq\ \mathbf{H}\left(\frac{r_i}{N};\ i\right) + \mathbf{H}\left(\frac{c_j}{N};\ j\right) \tag{5.4.1}$$

with the equality if and only if

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

and

$$\mathbf{H}\left(\frac{1-z_{ij}}{mn-N};\ i,j\right) \ \leq\ \mathbf{H}\left(\frac{n-r_i}{mn-N};\ i\right) + \mathbf{H}\left(\frac{m-c_j}{mn-N};\ j\right) \tag{5.4.2}$$

with the equality if and only if

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

Thus we have equalities in (5.4.1) and (5.4.2) if and only if

$$(r_i m - N)(c_j n - N) = 0 \quad \text{for all} \quad i,j,$$

that is, when all row sums are equal or all column sums are equal. In that case $I_0(R,C)$ estimates $|A_0(R,C)|$ within an $(mn)^{O(m+n)}$ factor. In all other cases, $I_0(R,C)$ overestimates $|A_0(R,C)|$ by as much as a $2^{\Omega(mn)}$ factor as long as the differences between the right hand sides and left hand sides of (5.4.1) and (5.4.2) multiplied by N and $(mn-N)$ respectively overcome the $O\big((m+n)\ln(mn)\big)$ error term, see also Section 5.5 for a particular family of examples.

We handle non-negative integer matrices slightly differently. For the independence estimate (5.1.2) we obtain

$$\begin{aligned}\ln I_+(R,C) = &-(N+mn)\mathbf{H}\left(\frac{r_i+n}{N+mn};\ i\right) - (N+mn)\mathbf{H}\left(\frac{c_j+m}{N+mn};\ j\right)\\ &- \sum_{i=1}^m r_i \ln r_i - \sum_{j=1}^n c_j \ln c_j\\ &+ N\ln N + (N+mn)\ln(N+mn) + O\big((m+n)\ln N\big)\end{aligned}$$

On the other hand, by Lemma 5.3 we have

$$\ln \alpha_+(R,C) = g\left(Z_+\right) \ \geq\ g(Y),$$

where Z_+ is the maximum entropy matrix and $Y=(y_{ij})$ is the matrix defined by

$$y_{ij} = \frac{r_i c_j}{N} \quad \text{for all} \quad i,j.$$

It is then easy to check that

$$\begin{aligned}g(Y) = &-(N+mn)\mathbf{H}\left(\frac{r_i c_j + N}{N(N+mn)};\ i,j\right)\\ &- \sum_{i=1}^m r_i \ln r_i - \sum_{j=1}^n c_j \ln c_j\\ &+ N\ln N + (N+mn)\ln(N+mn).\end{aligned}$$

By the inequality relating the entropy of a distribution and the entropy of its margins [Kh57], we have

$$(5.4.3) \qquad \mathbf{H}\left(\frac{r_i c_j + N}{N(N+mn)};\ i,j\right) \ \leq\ \mathbf{H}\left(\frac{r_i+n}{N+mn};\ i\right) + \mathbf{H}\left(\frac{c_j+m}{N+mn};\ j\right)$$

with the equality if and only if

$$\frac{r_i c_j + N}{N(N+mn)} = \frac{(r_i+n)(c_j+m)}{(N+mn)^2} \quad \text{for all} \quad i,j,$$

that is, when we have

$$(r_i m - N)(c_j n - N) = 0 \quad \text{for all} \quad i, j,$$

so that all row sums are equal or all column sums are equal. In that case, by symmetry we have $Y = Z_+$ and hence $I_+(R, C)$ estimates $|A_+(R, C)|$ within an $N^{O(m+n)}$ factor. In all other cases, $I_+(R, C)$ underestimates $|A_+(R, C)|$ by as much as a $2^{\Omega(mn)}$ factor as long as the difference between the right hand side and left hand side of (5.4.3) multiplied by $N + mn$ overcomes the $O\big((m+n)\ln N\big)$ error term, see also Section 5.5 for a particular family of examples.

(5.5) Cloning margins.

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

$$r_1 + \ldots + r_m = c_1 + \ldots + c_n = N.$$

For a positive integer k, let us define a km-vector R_k and a kn-vector C_k by

$$R_k = \Bigg(\underbrace{kr_1, \ldots, kr_1}_{k \text{ times}}, \ldots, \underbrace{kr_m, \ldots, kr_m}_{k \text{ times}} \Bigg) \quad \text{and}$$

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

We say that margins (R_k, C_k) are obtained by *cloning* from margins (R, C). It is not hard to show that if Z_0 and Z_+ are the maximum entropy matrices associated with margins (R, C) via Lemma 5.2 and Lemma 5.3 respectively, then the maximum entropy matrices associated with margins (R_k, C_k) are the Kronecker products $Z_0 \otimes Id_k$ and $Z_+ \otimes Id_k$ respectively, where Id_k is the $k \times k$ identity matrix. One has

$$\lim_{k \longrightarrow +\infty} |A_0(R_k, C_k)|^{1/k^2} = \alpha_0(R, C) \quad \text{and}$$

$$\lim_{k \longrightarrow +\infty} |A_+(R_k, C_k)|^{1/k^2} = \alpha_+(R, C).$$

Moreover, if not all coordinates r_i of R are equal and not all coordinates c_j of C are equal then the independence estimate $I_0(R_k, C_k)$, see (5.1.1), overestimates the number of $km \times kn$ matrices with row sums R_k and column sums C_k and 0-1 entries within a $2^{\Omega(k^2)}$ factor while the independence estimate $I_+(R_k, C_k)$, see (5.1.2), underestimates the number of $km \times kn$ non-negative integer matrices within a $2^{\Omega(k^2)}$ factor, see [Ba10a] and [Ba09] for details.

6. Random matrices with prescribed row and column sums

Estimates of Theorems 2.1 and 3.1, however crude, allow us to obtain a description of a random or typical matrix from sets $A_0(R,C)$ and $A_+(R,C)$, considered as finite probability spaces with the uniform measures.

Recall that x is a *Bernoulli* random variable if

$$\mathbf{Pr}\{x=0\}=p \quad \text{and} \quad \mathbf{Pr}\{x=1\}=q$$

for some $p,q \geq 0$ such that $p+q=1$. Clearly, $\mathbf{E}\, x = q$.

Recall that $P_0(R,C)$ is the polytope of $m \times n$ matrices with row sums R, column sums C and entries between 0 and 1. Let function $h: P_0(R,C) \longrightarrow \mathbb{R}$ and the maximum entropy matrix $Z_0 \in P_0(R,C)$ be defined as in Lemma 5.2.

The following result is proven in [Ba10a], see also [BH10a].

(6.1) Theorem. *Suppose that polytope $P_0(R,C)$ has a non-empty interior and let $Z_0 \in P_0(R,C)$ be the maximum entropy matrix. Let $X=(x_{ij})$ be a random $m \times n$ matrix of independent Bernoulli random variables x_{ij} such that $\mathbf{E}\, X = Z_0$. Then*

(1) *The probability mass function of X is constant on the set $A_0(R,C)$ of 0-1 matrices with row sums R and column sums C and*

$$\mathbf{Pr}\{X=D\} = e^{-h(Z_0)} \quad \textit{for all} \quad D \in A_0(R,C);$$

(2) *We have*

$$\mathbf{Pr}\ \{X \in A_0(R,C)\} \ \geq \ (mn)^{-\gamma(m+n)},$$

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

Theorem 6.1 implies that in many respects a random matrix $D \in A_0(R,C)$ behaves as a random matrix X of independent Bernoulli random variables such that $\mathbf{E}\, X = Z_0$, where Z_0 is the maximum entropy matrix. More precisely, any event that is sufficiently rare for the random matrix X (that is, an event the probability of which is essentially smaller than $(mn)^{-O(m+n)}$), will also be a rare event for a random matrix $D \in A_0(R,C)$. In particular, we can conclude that a typical matrix $D \in A_0(R,C)$ is sufficiently close to Z_0 as long as sums of entries over sufficiently large subsets S of indices are concerned.

For an $m \times n$ matrix $B=(b_{ij})$ and a subset

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

let

$$\sigma_S(B) = \sum_{(i,j)\in S} b_{ij}$$

be the sum of the entries of B indexed by set S. We obtain the following corollary, see [Ba10a] for details.

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

Let (R, C) be margins such that $n \geq m > q$ and the polytope $P_0(R, C)$ has a non-empty interior and let $Z_0 \in P_0(R, C)$ be the maximum entropy matrix. Let $S \subset \big\{(i, j) : \quad i = 1, \ldots, m; \ j = 1, \ldots, n\big\}$ be a set such that $\sigma_S(Z_0) \geq \delta mn$ and let

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

If $\epsilon \leq 1$ then

$$\mathbf{Pr}\Big\{D \in A_0(R, C) : \ (1 - \epsilon)\sigma_S(Z_0) \ \leq \ \sigma_S(D) \ \leq \ (1 + \epsilon)\sigma_S(Z_0)\Big\} \ \geq \ 1 - n^{-\kappa n}.$$

Recall that x is a *geometric* random variable if

$$\mathbf{Pr}\{x = k\} = pq^k \quad \text{for} \quad k = 0, 1, 2, \ldots$$

for some $p, q \geq 0$ such that $p + q = 1$. We have $\mathbf{E}\, x = q/p$.

Recall that $P_+(R, C)$ is the polytope of $m \times n$ non-negative matrices with row sums R and column sums C. Let function $g : P_+(R, C) \longrightarrow \mathbb{R}$ and the maximum entropy matrix $Z_+ \in P_0(R, C)$ be defined as in Lemma 5.3.

The following result is proven in [Ba10b], see also [BH10a].

(6.3) Theorem. *Let $Z_+ \in P_0(R, C)$ be the maximum entropy matrix. Let $X = (x_{ij})$ be a random $m \times n$ matrix of independent geometric random variables x_{ij} such that $\mathbf{E}\, X = Z_+$. Then*

1. *The probability mass function of X is constant on the set $A_+(R, C)$ of non-negative integer matrices with row sums R and column sums C and*

$$\mathbf{Pr}\{X = D\} = e^{-g(Z_+)} \quad \text{for all} \quad D \in A_+(R, C);$$

2. *We have*

$$\mathbf{Pr}\ \{X \in A_+(R, C)\} \ \geq \ N^{-\gamma(m+n)},$$

where $\gamma > 0$ is an absolute constant and $N = r_1 + \ldots + r_m = c_1 + \ldots + c_n$ for $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$.

Theorem 6.3 implies that in many respects a random matrix $D \in A_+(R, C)$ behaves as a matrix X of independent geometric random variables such that $\mathbf{E}\, X = Z_+$, where Z_+ is the maximum entropy matrix. More precisely, any event that is sufficiently rare for the random matrix X (that is, an event the probability of which is essentially smaller than $N^{-O(m+n)}$), will also be a rare event for a random matrix $D \in A_+(R, C)$. In particular, we can conclude that a typical matrix $D \in A_+(R, C)$ is sufficiently close to Z_+ as long as sums of entries over sufficiently large subsets S of indices are concerned.

Recall that $\sigma_S(B)$ denotes the sum of the entries of a matrix B indexed by a set S. We obtain the following corollary, see [Ba10b] for details.

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

Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_m)$ be positive integer vectors such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n = N$,

$$\begin{aligned} \frac{\delta N}{m} \ \leq \ r_i \ \leq \ \frac{N}{\delta m} \quad &for \quad i = 1, \ldots m, \\ \frac{\delta N}{n} \ \leq c_j \ \leq \ \frac{N}{\delta n} \quad &for \quad j = 1, \ldots, n \end{aligned}$$

and

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

Suppose that $n \geq m > q$ and let $S \subset \Big\{(i,j): \quad i = 1, \ldots, m, \ j = 1, \ldots, n\Big\}$ be a set such that $|S| \geq \delta mn$. Let $Z_+ \in P_+(R, C)$ be the maximum entropy matrix and let

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

If $\epsilon \leq 1$ then

$$\mathbf{Pr}\Big\{D \in A_+(R,C): \ (1-\epsilon)\sigma_S\left(Z_+\right) \ \leq \ \sigma_S(D) \ \leq \ (1+\epsilon)\sigma_S\left(Z_+\right)\Big\} \ \geq \ 1 - n^{-\kappa n}.$$

As is discussed in [BH10a], the ultimate reason why Theorems 6.1 and 6.3 hold true is that

the matrix X of independent Bernoulli random variables such that $\mathbf{E}\, X = Z_0$ is the random matrix with the maximum possible entropy among all random $m \times n$ matrices with 0-1 entries and the expectation in the affine subspace of the matrices with row sums R and column sums C

and

the matrix X of independent geometric random variables such that $\mathbf{E}\, X = Z_+$ is the random matrix with the maximum possible entropy among all random $m \times n$ matrices with non-negative integer entries and the expectation in the affine subspace of the matrices with row sums R and column sums C.

Thus Theorems 6.1 and 6.3 can be considered as an illustration of the Good's thesis [Go63] that the "null hypothesis" for an unknown probability distribution from a given class should be the hypothesis that the unknown distribution is, in fact, the distribution of the maximum entropy in the given class.

(6.5) Sketch of proof of Theorem 6.1. Let $Z_0 = (z_{ij})$ be the maximum entropy matrix as in Lemma 5.2. Let us choose $D \in A_0(R, C)$, $D = (d_{ij})$. Using (5.2.1), we get

$$\begin{aligned}
\mathbf{Pr}\left\{X = D\right\} =& \prod_{i,j} z_{ij}^{d_{ij}} \left(1 - z_{ij}\right)^{1-d_{ij}} = \prod_{ij} \frac{e^{(\lambda_i + \mu_j) d_{ij}}}{1 + e^{\lambda_i + \mu_j}} \\
=& \exp\left\{\sum_{i=1}^{m} \lambda_i r_i + \sum_{j=1}^{n} \mu_j c_j\right\} \prod_{ij} \frac{1}{1 + e^{\lambda_i + \mu_j}} \\
=& e^{-h(Z_0)},
\end{aligned}$$

which proves Part (1).

To prove Part (2), we use Part (1), Theorem 2.1 and Lemma 5.2. We have

$$\begin{aligned}
\mathbf{Pr}\left\{X \in A_0(R, C)\right\} =& \left|A_0(R, C)\right| e^{-h(Z_0)} \ \geq \ (mn)^{-\gamma(m+n)} \alpha_0(R, C) e^{-h(Z_0)} \\
=& (mn)^{-\gamma(m+n)}
\end{aligned}$$

for some absolute constant $\gamma > 0$. □

(6.6) Sketch of proof of Theorem 6.3. Let $Z_+ = (z_{ij})$ be the maximum entropy matrix as in Lemma 5.3. Let us choose $D \in A_+(R, C)$, $D = (d_{ij})$. Using (5.3.1), we get

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

which proves Part (1).

To prove Part (2), we use Part (1), Theorem 3.1 and Lemma 5.3. We have

$$\begin{aligned}
\mathbf{Pr}\left\{X \in A_+(R, C)\right\} =& \left|A_+(R, C)\right| e^{-g(Z_+)} \ \geq \ N^{-\gamma(m+n)} \alpha_+(R, C) e^{-g(Z_+)} \\
=& N^{-\gamma(m+n)}
\end{aligned}$$

for some absolute constant $\gamma > 0$. □

(6.7) Open questions. Theorems 6.1 and 6.3 show that a random matrix $D \in A_0(R, C)$, respectively $D \in A_+(R, C)$, in many respects behaves like a matrix of independent Bernoulli, respectively geometric, random variables whose expectation is the maximum entropy matrix Z_0, respectively Z_+. One can ask whether individual entries d_{ij} of D behave asymptotically as Bernoulli, respectively geometric, random variables with expectations z_{ij} as the size of the matrices grows. In the simplest situation we ask the following

(6.7.1) Question. Let (R, C) be margins and let (R_k, C_k) be margins obtained from (R, C) by cloning as in Section 5.5. Is it true that as k grows, the entry d_{11} of a random matrix $D \in A_0(R_k, C_k)$, respectively $D \in A_+(R_k, C_k)$, converges in distribution to the Bernoulli, respectively geometric, random variable with expectation z_{11}, where $Z_0 = (z_{ij})$, respectively $Z_+ = (z_{ij})$, is the maximum entropy matrix of margins (R, C)?

Some entries of the maximum entropy matrix Z_+ may turn out to be surprisingly large, even for reasonably looking margins. In [Ba10b], the following example is considered. Suppose that $m = n$ and let $R_n = C_n = (3n, n, \ldots, n)$. It turns out that the entry z_{11} of the maximum entropy matrix Z_+ is linear in n, namely $z_{11} > 0.58n$, while all other entries remain bounded by a constant. One can ask whether the d_{11} entry of a random matrix $D \in A_+(R_n, C_n)$ is indeed large, as the value of z_{11} suggests.

(6.7.2) Question. Let (R_n, C_n) be margins as above. Is it true that as n grows, one has $\mathbf{E}\, d_{11} = \Omega(n)$ for a random matrix $D \in A_+(R_n, C_n)$?

Curiously, the entry z_{11} becomes bounded by a constant if $3n$ is replaced by $2n$.

7. Asymptotic formulas for the number of matrices with prescribed row and column sums

In this section, we discuss asymptotically exact estimates for $|A_0(R, C)|$ and $|A_+(R, C)|$.

(7.1) An asymptotic formula for $|A_0(R, C)|$. Theorem 6.1 suggests the following way to estimate the number $|A_0(R, C)|$ of 0-1 matrices with row sums R and column sums C. Let us consider the matrix of independent Bernoulli random variables as in Theorem 6.1 and let Y be the random $(m+n)$-vector obtained by computing the row and column sums of X. Then, by Theorem 6.1, we have

$$|A_0(R, C)| = e^{h(Z_0)} \mathbf{Pr}\left\{X \in A_0(R, C)\right\} = e^{h(Z_0)} \mathbf{Pr}\left\{Y = (R, C)\right\}. \tag{7.1.1}$$

Now, random $(m+n)$-vector Y is obtained as a sum of mn independent random vectors and $\mathbf{E}\, Y = (R, C)$, so it is not unreasonable to assume that $\mathbf{Pr}\left\{Y = (R, C)\right\}$ can be estimated via some version of the Local Central Limit Theorem. In [BH10b] we show that this is indeed the case provided one employs the Edgeworth correction factor in the Central Limit Theorem.

We introduce the necessary objects to state the asymptotic formula for the number of 0-1 matrices with row sums R and column sums C.

Let $Z_0 = (z_{ij})$ be the maximum entropy matrix as in Lemma 5.2. We assume that $0 < z_{ij} < 1$ for all i and j. Let us consider the quadratic form $q_0 : \mathbb{R}^{m+n} \longrightarrow \mathbb{R}$ defined by

$$q_0(s, t) = \frac{1}{2} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} \left(z_{ij} - z_{ij}^2\right) \left(s_i + t_j\right)^2 \quad \text{for} \quad s = (s_1, \ldots, s_m) \quad \text{and} \quad t = (t_1, \ldots, t_n).$$

Quadratic form q_0 is positive semidefinite with the kernel spanned by vector

$$u = \Bigl(\underbrace{1, \ldots, 1}_{m \text{ times}}; \underbrace{-1, \ldots, -1}_{n \text{ times}} \Bigr).$$

Let $H = u^{\perp}$ be the hyperplane in $\mathbb{R}^{m+n}$ defined by the equation

$$s_1 + \ldots + s_m = t_1 + \ldots + t_n. \tag{7.1.2}$$

Then the restriction $q_0|H$ of q_0 onto H is a positive definite quadratic form and we define its determinant $\det q_0|H$ as the product of the non-zero eigenvalues of q_0. We consider the Gaussian probability measure on H with the density proportional to e^{-q_0} and define random variables $\phi_0, \psi_0 : H \longrightarrow \mathbb{R}$ by

$$\begin{aligned}\phi_0(s,t) =& \frac{1}{6} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} z_{ij} \left(1 - z_{ij}\right) \left(2 z_{ij} - 1\right) \left(s_i + t_j\right)^3 \quad \text{and} \\ \psi_0(s,t) =& \frac{1}{24} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} z_{ij} \left(1 - z_{ij}\right) \left(6 z_{ij}^2 - 6 z_{ij} + 1\right) \left(s_i + t_j\right)^4 \\ & \text{for} \quad (s,t) = \left(s_1, \ldots, s_m; t_1, \ldots, t_n\right).\end{aligned}$$

We let

$$\mu_0 = \mathbf{E}\, \phi_0^2 \quad \text{and} \quad \nu_0 = \mathbf{E}\, \psi_0.$$

(7.2) Theorem. *Let us fix $0 < \delta < 1/2$, let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be margins such that $m \geq \delta n$ and $n \geq \delta m$. Let $Z_0 = (z_{ij})$ be the maximum entropy matrix as in Lemma 5.2 and suppose that $\delta \leq z_{ij} \leq 1 - \delta$ for all i and j.*

Let the quadratic form q_0 and values μ_0 and ν_0 be as defined in Section 7.1. Then the number

$$\frac{e^{h(Z_0)} \sqrt{m+n}}{(4\pi)^{(m+n-1)/2} \sqrt{\det q_0|H}} \exp\left\{ -\frac{\mu_0}{2} + \nu_0 \right\} \tag{7.2.1}$$

approximates the number $|A_0(R,C)|$ of as $m, n \longrightarrow +\infty$ within a relative error which approaches 0 as $m, n \longrightarrow +\infty$. More precisely, for any $0 < \epsilon \leq 1/2$, the value of (7.2.1) approximates $|A_0(R,C)|$ within relative error ϵ provided

$$m, n \ \geq \ \left(\frac{1}{\epsilon} \right)^{\gamma(\delta)}$$

for some $\gamma(\delta) > 0$.

Some remarks are in order.

All the ingredients of formula (7.2.1) are efficiently computable, in time polynomial in $m+n$, see [BH10b] for details. If all row sums are equal then we have $z_{ij} = c_j/m$ by symmetry and if all column sums are equal, we have $z_{ij} = r_i/n$. In particular, if all row sums are equal and if all column sums are equal, we obtain the asymptotic formula of [C+08].

Let us consider formula (7.1.1). If, in the spirit of the Local Central Limit Theorem, we approximated $\mathbf{Pr}\left\{Y = (R,C)\right\}$ by $\mathbf{Pr}\left\{Y^* \in (R,C)+\Pi\right\}$, where Y^* is the $(m+n-1)$-dimensional random Gaussian vector whose expectation and covariance matrix match those of Y and where Π is the set of points on the hyperplane H that are closer to (R,C) than to any other integer vector in H, we would have obtained the first part

$$\frac{e^{h(Z_0)}\sqrt{m+n}}{(4\pi)^{(m+n-1)/2}\sqrt{\det q_0|H}}$$

of formula (7.2.1). Under the conditions of Theorem 7.2 we have

$$c_1(\delta) \ \leq \ \exp\left\{-\frac{\mu_0}{2}+\nu_0\right\} \ \leq \ c_2(\delta)$$

for some constants $c_1(\delta), c_2(\delta) > 0$ and this factor represents the Edgeworth correction to the Central Limit Theorem. We note that the constraints $\delta \leq z_{ij} \leq 1-\delta$ are, generally speaking, unavoidable. If the entries z_{ij} of the maximum entropy matrix are uniformly small, then the distribution of the random vector Y of row and column sums of the random Bernoulli matrix X is no longer approximately Gaussian but approximately Poisson and formula (7.2.1) does not give correct asymptotics. The sparse case of small row and column sums is investigated in [G+06].

More generally, to have some analytic formula approximating $|A_0(R,C)|$ we need certain regularity conditions on (R,C), since the number $|A_0(R,C)|$ becomes volatile when the margins (R,C) approach the boundary of the Gale-Ryser conditions, cf. [JSM92]. By requiring that the entries of maximum entropy matrix Z_0 are separated from both 0 and 1, we ensure that the margins (R,C) remain sufficiently inside the polyhedron defined by the Gale-Ryser inequality and the number of 0-1 matrices with row sums R and column sums C changes sufficiently smoothly when R and C change.

(7.3) An asymptotic formula for $|A_+(R,C)|$. As in Theorem 6.3, let X be the matrix of independent geometric random variables such that $\mathbf{E}\, X = Z_+$, where Z_+ is the maximum entropy matrix. Let Y be the random $(m+n)$-vector obtained by computing the row and column sums of X. Then, by Theorem 6.3, we have

$$(7.3.1) \qquad |A_+(R,C)| = e^{g(Z_+)}\mathbf{Pr}\left\{X \in A_+(R,C)\right\} = e^{g(Z_+)}\mathbf{Pr}\left\{Y = (R,C)\right\}.$$

In [BH09] we show how to estimate the probability that $Y=(R,C)$ using the Local Central Limit Theorem with the Edgeworth correction.

Let $Z_+ = (z_{ij})$ be the maximum entropy matrix as in Lemma 5.3. Let us consider the quadratic form $q_+ : \mathbb{R}^{m+n} \longrightarrow \mathbb{R}$ defined by

$$q_+(s,t) = \frac{1}{2} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} \left(z_{ij} + z_{ij}^2\right) \left(s_i + t_j\right)^2$$
$$\text{for} \quad s = (s_1, \dots, s_m) \quad \text{and} \quad t = (t_1, \dots, t_n).$$

Let $H \subset \mathbb{R}^{m+n}$ be the hyperplane defined by (7.1.2). The restriction $q_+|H$ of q_+ onto H is a positive definite quadratic form and we define its determinant $\det q_+|H$ as the product of the non-zero eigenvalues of q_+. We consider the Gaussian probability measure on H with the density proportional to e^{-q_+} and define random variables $\phi_+, \psi_+ : H \longrightarrow \mathbb{R}$ by

$$\phi_+(s,t) = \frac{1}{6} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} z_{ij} \left(1 + z_{ij}\right) \left(2 z_{ij} + 1\right) \left(s_i + t_j\right)^3 \quad \text{and}$$
$$\psi_+(s,t) = \frac{1}{24} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} z_{ij} \left(1 + z_{ij}\right) \left(6 z_{ij}^2 + 6 z_{ij} + 1\right) \left(s_i + t_j\right)^4$$
$$\text{for} \quad (s,t) = (s_1, \dots, s_m; t_1, \dots, t_n).$$

We let

$$\mu_+ = \mathbf{E}\, \phi_+^2 \quad \text{and} \quad \nu_+ = \mathbf{E}\, \psi_+.$$

(7.4) Theorem. *Let us fix $0 < \delta < 1$, let $R = (r_1, \dots, r_m)$ and $C = (c_1, \dots, c_n)$ be margins such that $m \ge \delta n$ and $n \ge \delta m$. Let $Z_+ = (z_{ij})$ be the maximum entropy matrix as in Lemma 5.3. Suppose that*

$$\delta \tau \ \le\ z_{ij} \ \le\ \tau \quad \textit{for all} \quad i, j$$

for some $\tau \ge \delta$.

Let the quadratic form q_+ and values μ_+ and ν_+ be as defined in Section 7.3. Then the number

$$\frac{e^{g(Z_+)} \sqrt{m+n}}{(4\pi)^{(m+n-1)/2} \sqrt{\det q_+|H}} \exp\left\{ -\frac{\mu_+}{2} + \nu_+ \right\} \tag{7.4.1}$$

approximates the number $|A_+(R,C)|$ of as $m, n \longrightarrow +\infty$ within a relative error which approaches 0 as $m, n \longrightarrow +\infty$. More precisely, for any $0 < \epsilon \le 1/2$, the value of (7.4.1) approximates $|A_+(R,C)|$ within relative error ϵ provided

$$m, n \ \ge\ \left(\frac{1}{\epsilon}\right)^{\gamma(\delta)}$$

for some $\gamma(\delta) > 0$.

All the ingredients of formula (7.4.1) are efficiently computable, in time polynomial in $m+n$, see [BH09] for details. If all row sums are equal then we have $z_{ij} = c_j/m$ by symmetry and if all column sums are equal, we have $z_{ij} = r_i/n$. In particular, if all row sums are equal and if all column sums are equal, we obtain the asymptotic formula of [C+07]. The term

$$\frac{e^{g(Z_+)}\sqrt{m+n}}{(4\pi)^{(m+n-1)/2}\sqrt{\det q_+|H}}$$

corresponds to the Gaussian approximation for the distribution of the random vector Y in (7.3.1), while

$$\exp\left\{-\frac{\mu_+}{2}+\nu_+\right\}$$

is the Edgeworth correction factor.

While the requirement that the entries of the maximum entropy matrix Z_+ are separated from 0 is unavoidable (if z_{ij} are small, the coordinates of Y are asymptotically Poisson, not Gaussian, see [GM08] for the analysis of the sparse case), it is not clear whether the requirement that all z_{ij} are within a constant factor of each other is indeed needed. It could be that around certain margins (R, C) the number $|A_+(R, C)|$ experiences sudden jumps, as the margins change, which precludes the existence of an analytic expression similar to (7.4.1) for $|A_+(R, C)|$. A candidate for such an abnormal behavior is supplied by the margins discussed in Section 6.7. Namely, if $m = n$ and $R = C = (\lambda n, n, \ldots, n)$ then for $\lambda = 2$ all the entries of the maximum entropy matrix Z_+ are $O(1)$, while for $\lambda = 3$ the first entry z_{11} grows linearly in n. Hence for some particular λ between 2 and 3 a certain "phase transition"' occurs: the entry z_{11} jumps from $O(1)$ to $\Omega(n)$. It would be interesting to find out if there is indeed a sharp change in $|A_+(R, C)|$ when λ changes from 2 to 3.

8. Concluding remarks

Method of Sections 6 and 7 have been applied to some related problems, such as counting higher-order "tensors" with 0-1 or non-negative integer entries and prescribed sums along coordinate hyperplanes [BH10a] and counting graphs with prescribed degrees of vertices [BH10b], which corresponds to counting symmetric 0-1 matrices with zero trace and prescribed row (column) sums.

In general, the problem can be described as follows: we have a polytope $P \subset \mathbb{R}^d$ defined as the intersection of the non-negative orthant $\mathbb{R}^d_+$ with an affine subspace $\mathcal{A}$ in $\mathbb{R}^d$ and we construct a d-vector X of independent Bernoulli (in the 0-1 case) or geometric (in the non-negative integer case) random variables, so that the expectation of X lies in $\mathcal{A}$ and the distribution of X is uniform, when restricted onto the set of 0-1 or integer points in P. Random vector X is determined by its expectation $\mathbf{E}\, X = z$ and z is found by solving a convex optimization problem on

P. Since vector X conditioned on the set of 0-1 or non-negative integer vectors in P is uniform, the number of 0-1 or non-negative integer points in P is expressed in terms of the probability that X lies in $\mathcal{A}$. Assuming that the affine subspace $\mathcal{A}$ is defined by a system $Ax = b$ of linear equations, where A is $k \times d$ matrix of rank $k < d$, we define a k-vector $Y = AX$ of random variables and estimate the probability that $Y = b$ by using a Local Central Limit Theorem type argument. Here we essentially use that $\mathbf{E}\, Y = b$, since the expectation of X lies in $\mathcal{A}$.

Not surprisingly, the argument works the easiest when the codimension k of the affine subspace (and hence the dimension of vector Y) is small. In particular, counting higher-order "tensors" is easier than counting matrices, the need in the Edgeworth correction factor, for example, disappears as the vector Y turns out to be closer in distribution to a Gaussian vector, see [BH10a]. Once a Gaussian or almost Gaussian estimate for the probability $\mathbf{Pr}\left\{Y = b\right\}$ is established, one can claim a certain concentration of a random 0-1 or integer point in P around $z = \mathbf{E}\, X$.

References

[Ba07] A. Barvinok, *Brunn-Minkowski inequalities for contingency tables and integer flows*, Advances in Mathematics **211** (2007), 105–122.

[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.

[Ba10a] A. Barvinok, *On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries*, Advances in Mathematics **224** (2010), 316–339.

[Ba10b] A. Barvinok, *What does a random contingency table look like?*, Combinatorics, Probability and Computing **19** (2010), 517–539.

[BH09] A. Barvinok and J.A. Hartigan, *An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sum*, preprint `arXiv:0910.2477` (2009).

[BH10a] A. Barvinok and J.A. Hartigan, *Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes*, Advances in Applied Mathematics **45** (2010), 252–289.

[BH10b] A. Barvinok and J.A. Hartigan, *The number of graphs and a random graph with a given degree sequence*, preprint `arXiv:1003.0356` (2010).

[Be74] E. Bender, *The asymptotic number of non-negative integer matrices with given row and column sums*, Discrete Math. **10** (1974), 217–223.

[BR91] R.A. Brualdi and H.J. Ryser, *Combinatorial Matrix Theory*, Encyclopedia of Mathematics and its Applications, **39**, Cambridge University Press, Cambridge, 1991.

[C+08] E.R. Canfield, C. Greenhill, and B.D. McKay, *Asymptotic enumeration of dense 0-1 matrices with specified line sums*, Journal of Combinatorial Theory. Series A **115** (2008), 32–66.

[C+07] E.R. Canfield and B.D. McKay, *Asymptotic enumeration of contingency tables with constant margins*, preprint `arXiv` math.CO/0703600, Combinatorica, to appear (2007).

[Ga02] R.J. Gardner, *The Brunn-Minkowski inequality*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), 355–405.

[Go63] I.J. Good, *Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables*, Ann. Math. Statist. **34** (1963), 911–934.

[Go76] I.J. Good, *On the application of symmetric Dirichlet distributions and their mixtures to contingency tables*, Ann. Statist. **4** (1976), 1159–1189.

[GC77] I.J. Good and J.F. Crook, *The enumeration of arrays and a generalization related to contingency tables*, Discrete Mathematics **19** (1977), 23–45.

[G+06] C. Greenhill, B.D. McKay, and X. Wang, *Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums*, Journal of Combinatorial Theory. Series A **113** (2006), 291–324.

[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), 59–481.

[JSM92] M. Jerrum, A. Sinclair and B. McKay, *When is a graphical sequence stable?*, Random Graphs, Vol. 2 (Poznań, 1989), Wiley-Intersci. Publ., Wiley, New York, 1992, pp. 101–115.

[Kh57] A.I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications Inc., New York, N. Y., 1957.

[LW01] J.H. van Lint and R.M. Wilson, *A Course in Combinatorics. Second edition*, Cambridge University Press, Cambridge, 2001.

[NN94] Y. 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