ASYMPTOTIC ESTIMATES FOR THE NUMBER OF CONTINGENCY TABLES, INTEGER FLOWS, AND VOLUMES OF TRANSPORTATION POLYTOPES

Alexander Barvinok

August 2008

Abstract. We prove an asymptotic estimate for the number of $m \times n$ non-negative integer matrices (contingency tables) with prescribed row and column sums and, more generally, for the number of integer feasible flows in a network. Similarly, we estimate the volume of the polytope of $m \times n$ non-negative real matrices with prescribed row and column sums. Our estimates are solutions of convex optimization problems and hence can be computed efficiently. As a corollary, we show that if row sums $R = (r_1, \ldots, r_m)$ and column sums $C = (c_1, \ldots, c_n)$ with $r_1 + \ldots + r_m = c_1 + \ldots + c_n = N$ are sufficiently far from constant vectors, then, asymptotically, in the uniform probability space of the $m \times n$ non-negative integer matrices with the total sum N of entries, the event consisting of the matrices with row sums R and the event consisting of the matrices with column sums C are positively correlated.

1. Introduction and main results

Let $m > 1$ and $n > 1$ be integers and let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive integer vectors such that

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

We are interested in the number $\#(R, C)$ of $m \times n$ non-negative integer matrices, also known as *contingency tables*, with row sums R and column sums C, called *margins*. Computing or estimating numbers $\#(R, C)$ has attracted a lot of attention, because of the relevance of these numbers in statistics, see [Goo76], [DE85], combinatorics, representation theory, and elsewhere, see [DG85], [DG04]. Of interest

1991 *Mathematics Subject Classification.* 05A16, 60C05, 52A38, 52B12, 52B55.

Key words and phrases. contingency tables, transportation polytope, volume estimates, asymptotic estimates, integer flows.

This research was partially supported by NSF Grant DMS 0400617.

are asymptotic formulas, see [BBK72], [Ben74] and most recent [CM07a], [GM07], algorithms with rigorous estimates of the performance guarantees, see [DKM97], [Mor02], [CD03], [BLV04], and heuristic approaches which may lack formal justification but tend to work well in practice [Goo76], [DE85], [C+05].

Our first main result is as follows.

(1.1) 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

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

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

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

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

More precisely, the lower bound we prove is

$$\#(R, C) \ \geq \ \frac{\Gamma\left(\frac{m+n}{2}\right)}{2e^5 \pi^{\frac{m+n-2}{2}} mn(N + mn)} \left(\frac{2}{(mn)^2 (N+1)(N+mn)}\right)^{m+n-1}$$
$$\times \left(\prod_{i=1}^{m} \frac{r_i^{r_i}}{r_i!}\right) \left(\prod_{j=1}^{n} \frac{c_j^{c_j}}{c_j!}\right) \frac{N!(N+mn)!(mn)^{mn}}{N^N (N+mn)^{N+mn} (mn)!} \rho(R, C)$$

provided $m + n \geq 10$. Recall that from Stirling's formula

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

and hence the product in front of $\rho(R, C)$ indeed exceeds $N^{-\gamma(m+n)}$ for some absolute constant $\gamma > 0$.

We note that the substitution $x_i = e^{-t_i}$ and $y_j = e^{-s_j}$ transforms the problem of computing ρ into the problem of minimizing the convex function

$$\phi(\mathbf{t}, \mathbf{s}) = \phi_{R,C}(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^{m} r_i t_i + \sum_{j=1}^{n} c_j s_j - \sum_{ij} \ln\left(1 - e^{-t_i - s_j}\right)$$

on the positive orthant $s_i, t_j > 0$, so that methods of convex optimization can be applied to compute ρ in time polynomial in $m+n$ and $\ln N$, see [NN94].

Theorem 1.1 estimates the number $\#(R,C)$ of contingency tables within an $N^{O(m+n)}$ factor. This estimate provides, asymptotically, the main term of $\log \#(R,C)$ for all but very sparse cases, where margins r_i and c_j are small compared to the sizes m and n of the matrix. For example, if the margins r_i and c_j are at least linear in m and n then $\#(R,C)$ is at least as big as γ^{mn} for some constant $\gamma > 1$. By now, the sparse case of small r_i and c_j is well understood, thanks especially to the recent paper [GM07]. The case of moderate to high margins seems to be the most difficult. To the author's knowledge, the estimate of Theorem 1.1 is the *only* rigorously proven effective estimate of $\#(R,C)$ for generic R and C (if all r_i's are equal and all c_j's are equal, recent paper [CM07a] provides a precise asymptotic formula for the number of tables). Theorem 1.1 allows us to find faults with the very intuitive "independence heuristic" for counting contingency tables and points out at some strange "attraction" phenomena in the space of matrices. Quite counter-intuitively, we conclude that in the uniform probability space of the $m \times n$ non-negative integer matrices with the total sum of entries equal to N, the event consisting of the matrices with row sums R and the event consisting of the matrices with column sums C attract exponentially in mn provided the vectors R and C are sufficiently far from constant vectors, see Section 2 for the precise statements and details.

Let us identify the space of $m \times n$ real matrices $X = (x_{ij})$ with Euclidean space $\mathbb{R}^d$ for $d = mn$. In $\mathbb{R}^d$ we consider the *transportation polytope* $\mathcal{P} = \mathcal{P}(R,C)$ defined by the equations

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

and inequalities

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

As is known, $\mathcal{P}$ is a polytope of dimension $(m-1)(n-1)$. We prove the following estimate for the volume of $\mathcal{P}$, computed with respect to the Euclidean structure in the affine span of $\mathcal{P}$, induced from $\mathbb{R}^d$.

(1.2) Theorem. *Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive integer vectors such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n = N$ and let $\mathcal{P} = \mathcal{P}(R,C)$ be the polytope of non-negative $m \times n$ matrices with row sums $r_1, \ldots, r_m$ and column sums $c_1, \ldots, c_n$.*

Let

$$\beta = \beta(R,C) = \max_{\substack{X=(x_{ij}) \\ X \in \mathcal{P}}} \prod_{ij} x_{ij}$$

be the maximum value of the product of entries of a matrix from $\mathcal{P}$. Then for the volume of $\mathcal{P}$ we have

$$\beta e^{mn} N^{\gamma(m+n)} \ \geq \ \operatorname{vol}(\mathcal{P}) \ \geq \ \beta e^{mn} N^{-\gamma(m+n)},$$

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

From our proof more precise bounds

$$\begin{aligned}\operatorname{vol}(\mathcal{P}) \ &\geq \ \frac{\Gamma\left(\frac{m+n}{2}\right)}{2e^3\sqrt{mn}\pi^{\frac{m+n-2}{2}}N^{m+n-1}}\frac{(mn)^{mn}}{(mn)!}\beta \\ &\text{and} \\ \operatorname{vol}(\mathcal{P}) \ &\leq \ \frac{2e\lambda^{m+n-2}(mn)^{2m+2n-5/2}}{N^{m+n-1}}\frac{(mn)^{mn}}{(mn)!}\beta, \\ &\text{where}\end{aligned}$$

$$\lambda = \lambda(R, C) = \frac{n}{2}\max_{i=1,\dots,m}\frac{N}{r_i} + \frac{m}{2}\max_{j=1,\dots,n}\frac{N}{c_j}$$

follow. When the margins are scaled, $(R, C) \longmapsto (tR, tC)$ for $t > 0$, the volume of $\mathcal{P}$ and both the upper and the lower bounds get multiplied by $t^{\dim \mathcal{P}}$.

Computing β reduces to finding the maximum of the concave function

$$f(X) = \sum_{ij} \ln x_{ij}$$

on the transportation polytope $\mathcal{P}$ and hence can be done efficiently (in time polynomial in $m+n$ and $\ln N$) by existing methods [NN94].

Computing or estimating volumes of transportation polytopes has attracted considerable attention as a testing ground for methods of convex geometry [Sch92], combinatorics [Pak00], analysis and algebra [BLV04], [BP03], [DLY03]. In a recent breakthrough [CM07b], Canfield and McKay obtained a precise asymptotic expression for the volume of the *Birkhoff polytope* (when $r_i = c_j = 1$ for all i and j) and in the more general case of all the row sums being equal and all the column sums being equal. If $R = C = (1, \dots, 1)$, the formula of [CM07b] gives

$$\operatorname{vol}\mathcal{P} = \frac{1}{(2\pi)^{n-1/2}n^{(n-1)^2}}\exp\left\{\frac{1}{3} + n^2 + O\left(n^{-1/2+\epsilon}\right)\right\},$$

whereas the formula of Theorem 1.2 implies that, ignoring lower-order terms, we have

$$\operatorname{vol}\mathcal{P} \approx \frac{e^{n^2}}{n^{n^2}}$$

in that case (since by symmetry the maximum β of the product of coordinates x_{ij} is attained at $x_{ij} = 1/n$). Theorem 1.2 seems to be the only rigorously proven estimate of the volume of the transportation polytope available for general margins.

We note that from the purely algorithmic perspective, volumes of polytopes and convex bodies can be computed in *randomized* polynomial time, see [Bol97] for a survey.

Theorem 1.1 can be extended to counting with weights.

Let us fix a non-negative matrix $W = (w_{ij})$, which we call the matrix of *weights*. We consider the following expression

$$T(R, C; W) = \sum_{D=(d_{ij})} \prod_{ij} w_{ij}^{d_{ij}},$$

where the sum is taken over all non-negative integer matrices D with row sums R and column sums C and where we agree that $0^0 = 1$. For example, if $w_{ij} \in \{0, 1\}$ for all i, j then $T(R, C; W)$ is the number of $m \times n$ non-negative integer matrices $D = (d_{ij})$ with row sums R, column sums C and such that $d_{ij} = 0$ whenever $w_{ij} = 0$. This number can also be interpreted as the number of integer feasible flows in a bipartite graph with vertices $u_1, \ldots, u_m$ and $v_1, \ldots, v_n$ and edges (u_i, v_j) whenever $w_{ij} = 1$ that satisfy the supply constraints r_i at u_i and the demand constraints c_j at v_j. Counting integer feasible flows in non-bipartite networks can be reduced to that for bipartite networks. For example, if $w_{ij} = 1$ for $j \leq i+1$ and $w_{ij} = 0$ elsewhere, $T(R, C; W)$ is the Kostant partition function, see [Ba07], [Ba08] for more examples and details. We also note that $T(R, C; \mathbf{1}) = \#(R, C)$, where $\mathbf{1}$ is the matrix of all 1's.

We prove the following extension of Theorem 1.1.

(1.3) Theorem. *Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive integer vectors such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n = N$ and let $W = (w_{ij})$ be an $m \times n$ non-negative matrix of weights. Let us define a function*

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

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

and let

$$\rho = \rho(R, C; W) = \inf_{\substack{x_1, \ldots, x_m > 0 \\ y_1, \ldots, y_n > 0 \\ w_{ij} x_i y_j < 1 \text{ for all } i,j}} F(\mathbf{x}, \mathbf{y}; W).$$

Then, for the number $T(R, C; W)$ of weighted non-negative integer matrices with row sums $r_1, \ldots, r_m$ and column sums $c_1, \ldots, c_n$, we have

$$\rho \ \geq \ T(R, C; W) \ \geq \ N^{-\gamma(m+n)} \rho,$$

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

More precisely, the lower bound we prove is

$$T(R, C; W) \ \geq \ \frac{\Gamma\left(\frac{m+n}{2}\right)}{2e^5 \pi^{\frac{m+n-2}{2}} mn(N + mn)} \left(\frac{2}{(mn)^2 (N+1)(N+mn)}\right)^{m+n-1}$$
$$\times \left(\prod_{i=1}^m \frac{r_i^{r_i}}{r_i!}\right)\left(\prod_{j=1}^n \frac{c_j^{c_j}}{c_j!}\right) \frac{N!(N+mn)!(mn)^{mn}}{N^N (N+mn)^{N+mn} (mn)!} \rho(R, C; W)$$

provided $m+n \geq 10$.

As in Theorem 1.1, substituting $x_i = e^{-t_i}$ for $i=1, \ldots, m$ and $y_j = e^{-s_j}$ for $j=1, \ldots, n$ we reduce the problem of computing ρ to the problem of finding the infimum of the convex function

$$\phi(\mathbf{t}, \mathbf{s}) = \phi_{R,C}(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^{m} r_i t_i + \sum_{j=1}^{n} c_j s_j - \sum_{ij} \ln\left(1 - w_{ij} e^{-t_i - s_j}\right)$$

on the convex polyhedron

$$s_i + t_j > \ln w_{ij} \quad \text{for all} \quad i, j.$$

Again, the value of ρ can be computed efficiently, both in theory and in practice, by methods of convex optimization, cf. [NN94].

For *positive* matrices $W = (w_{ij})$ the infimum $\rho(R, C; W)$ in Theorem 1.3 is attained at a particular point and there is a convenient dual description of $\rho(R, C; W)$.

(1.4) Lemma. *Let $\mathcal{P} = \mathcal{P}(R, C)$ be the transportation polytope of the $m \times n$ non-negative matrices $X = (x_{ij})$ with row sums R and column sums C and let us fix an $m \times n$ positive matrix $W = (w_{ij})$ of weights, so $w_{ij} > 0$ for all i, j. For an $m \times n$ non-negative matrix $X = (x_{ij})$ let us define*

$$g(X; W) = \sum_{ij} \Big((x_{ij} + 1) \ln (x_{ij} + 1) - x_{ij} \ln x_{ij} + x_{ij} \ln w_{ij} \Big).$$

Then $g(X; W)$ is a strictly concave function of X and attains its maximum on $\mathcal{P}$ at a unique positive matrix $Z = Z(R, C; W)$. One can write $Z = (z_{ij})$ in the form

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

and positive $\xi_1, \ldots, \xi_m; \eta_1, \ldots, \eta_n$ such that $w_{ij} \xi_i \eta_j < 1$ for all i and j and such that the infimum $\rho(R, C; W)$ in Theorem 1.3 is attained at $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$:

$$\rho(R, C; W) = F\left(\mathbf{x}^*, \mathbf{y}^*; W\right).$$

Moreover, we have

$$\rho(R, C; W) = \exp\Big\{ g(Z; W) \Big\}.$$

In particular, if $w_{ij} = 1$ for all i, j, then

$$g(X) = g(X; \mathbf{1}) = \sum_{ij} \Big((x_{ij} + 1) \ln (x_{ij} + 1) - x_{ij} \ln x_{ij} \Big) \quad \textit{and}$$

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

where $\mathbf{x}^* = (\xi_1, \dots, \xi_m)$ *and* $\mathbf{y}^* = (\eta_1, \dots, \eta_n)$ *is a point where the minimum* $\rho(R,C) = F(\mathbf{x}^*, \mathbf{y}^*)$ *in Theorem 1.1 is attained. Additionally,*

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

The paper is structured as follows.

In Section 2, we consider consequences of Theorems 1.1 and 1.2 for the "independence heuristic". The heuristic was, apparently, first discussed by Good, see [Goo76]. It asserts that if we consider the space of non-negative integer $m \times n$ matrices with the total sum N of entries as a probability space with the uniform measure then the event consisting of the matrices with the row sums $r_1, \dots, r_m$ is "almost independent" from the event consisting of the matrices with the column sums $c_1, \dots, c_n$. We show that if the row sums r_i and the column sums c_j are sufficiently generic then the independence heuristic tends to underestimate the number of tables as badly as within a factor of γ^{mn} for some absolute constant $\gamma > 1$. We see that in fact (rather counter-intuitively), instead of independence, we have attraction (positive correlation) of the events.

In Section 3, we state a general result (Theorem 3.1), which provides a reasonably accurate estimate for the volume of the section of the standard simplex by a subspace of a small codimension. Theorem 3.1 states that in a sufficiently generic situation the volume of the section is determined by the maximum value of the product of the coordinates of a point in the section. This estimate immediately implies Theorem 1.2 and is one of the two crucial ingredients in the proofs of Theorems 1.1 and 1.3. Theorem 3.1 appears to be new and may be interesting in its own right.

In Section 4, we state some preliminaries from convex geometry needed to prove Theorem 3.1.

In Section 5, we prove Theorems 3.1 and 1.2.

In Section 6, we describe the second main ingredient for the proofs of Theorems 1.1 and 1.3, the integral representation from [Ba07] and [Ba08] for the number $\#(R,C)$ of tables and the number $T(R,C;W)$ of weighted tables.

In Section 7, we prove Theorems 1.1 and 1.3 and Lemma 1.4.

In what follows, we use γ to denote a positive constant.

2. The independence heuristic and the exponential attraction in the space of matrices

(2.1) The independence heuristic. The following heuristic approach to counting contingency tables was suggested by Good [Goo76]. Let us consider the space of all $m \times n$ non-negative integer matrices with the total sum of entries N as a probability space with the uniform measure. Then the probability that a matrix from this space has row sums $R=(r_1, \ldots, r_m)$ is exactly

$$\binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{r_i+n-1}{n-1}.$$

Similarly, the probability that a matrix has column sums $C=(c_1, \ldots, c_n)$ is exactly

$$\binom{N+mn-1}{mn-1}^{-1} \prod_{j=1}^{n} \binom{c_j+m-1}{m-1}.$$

Assuming that the two events are almost independent, one estimates the number $\#(R,C)$ of contingency tables by the *independence heuristic* $I(R,C)$:

$$I(R,C) = \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{r_i+n-1}{n-1} \prod_{j=1}^{n} \binom{c_j+m-1}{m-1}. \tag{2.1.1}$$

For example, if $m=n=4$, $R=(220, 215, 93, 64)$, $C=(108, 286, 71, 127)$ with $N=592$ then

$$\#(R,C) = 1225914276768514 \approx 1.226 \times 10^{15},$$

see [DE85], while

$$I(R,C) \approx 1.211 \times 10^{15}.$$

Given margins $R=(r_1, \ldots, r_n)$ and $C=(c_1, \ldots, c_m)$ such that not all row sums r_i are equal and not all column sums c_j are equal, we will construct a sequence of margins (R_k, C_k), where R_k is a km-vector and C_k is a kn-vector such that the ratio $\#(R_k, C_k)/I(R_k, C_k)$ grows as γ^{k^2} for some $\gamma = \gamma(R,C) > 1$.

(2.2) Cloning margins. Let us choose some margins $R=(r_1, \ldots, r_m)$ and $C=(c_1, \ldots, c_n)$ such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n = N$. For a positive integer k, let us consider the new "clone" margins

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

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

In other words, we obtain margins (R_k, C_k) if we choose an arbitrary matrix X with row sums R and column sums C, consider the $km \times kn$ block matrix Y_k consisting of k^2 blocks X and let R_k be the row sums of Y_k and let C_k be the column sums of Y_k. Hence we consider $km \times kn$ matrices with the total sum of the matrix entries equal to $k^2 N$.

One can check from the optimality condition (cf. Section 7.1) that if $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$ is a point in Theorem 1.1 where the minimum $\rho(R, C)$ is attained then the minimum $\rho(R_k, C_k)$ is attained at the point

$$\Bigl(\underbrace{\mathbf{x}^*, \ldots, \mathbf{x}^*}_{k \text{ times}}, \underbrace{\mathbf{y}^*, \ldots, \mathbf{y}^*}_{k \text{ times}}\Bigr).$$

Therefore,

$$\rho(R_k, C_k) = \rho^{k^2}(R, C)$$

and by Theorem 1.1

$$\lim_{k \longrightarrow +\infty} \#(R_k, C_k)^{1/k^2} = \rho(R, C),$$

or, in other words,

$$\lim_{k \longrightarrow +\infty} \frac{1}{k^2} \ln \#(R_k, C_k) = \ln \rho(R, C). \tag{2.2.1}$$

Let us introduce the multivariate entropy function

$$\mathbf{H}(p_1, \ldots, p_d) = \sum_{i=1}^{d} p_i \ln \frac{1}{p_i},$$

where $p_1, \ldots, p_d$ are non-negative numbers such that $p_1 + \ldots + p_d = 1$. Using the standard asymptotic estimate for binomial coefficients (available, for example, via Stirling's formula)

$$\lim_{k \longrightarrow +\infty} \frac{1}{k} \ln \binom{ka + kb}{ka} = (a + b) \ln(a + b) - a \ln a - b \ln b$$

we deduce from (2.1.1) that

$$\begin{aligned} \lim_{k \longrightarrow +\infty} \frac{1}{k^2} \ln I(R_k, C_k) = & \\ & - (N + mn)\mathbf{H}\left(\frac{r_i + n}{N + mn},\ i = 1, \ldots, m\right) \\ & - (N + mn)\mathbf{H}\left(\frac{c_j + m}{N + mn},\ j = 1, \ldots, n\right) \\ & - \sum_{i=1}^{m} r_i \ln r_i - \sum_{j=1}^{n} c_j \ln c_j \\ & + N \ln N + (N + mn) \ln(N + mn) \end{aligned} \tag{2.2.2}$$

(2.3) The exponential attraction in the space of matrices. Let us choose margins $R=(r_1, \dots, r_m)$ and $C=(c_1, \dots, c_n)$ such that not all row sums r_i are equal and not all column sums c_j are equal. Our goal is to show that

$$\lim_{k \longrightarrow +\infty} \frac{1}{k^2} \ln \#(R_k, C_k) \ > \ \lim_{k \longrightarrow +\infty} \frac{1}{k^2} \ln I(R_k, C_k), \tag{2.3.1}$$

so the ratio $\#(R_k, C_k)/I(R_k, C_k)$ grows as γ^{k^2} for some $\gamma=\gamma(R,C)>1$, as we clone margins $(R,C) \longmapsto (R_k, C_k)$.

By Lemma 1.4, we can write

$$\ln \rho(R,C) \ = \ g(Z) \ \geq \ g(Y), \tag{2.3.2}$$

where $Y=(y_{ij})$ is the *independence matrix* with $y_{ij}=r_i c_j/N$ for all i,j and

$$g(X)=\sum_{ij} \Big((x_{ij}+1) \ln (x_{ij}+1) - x_{ij} \ln x_{ij} \Big).$$

On the other hand, it is easy to check that

$$\begin{aligned} g(Y) = - (N+mn) \mathbf{H} \left(\frac{r_i c_j + N}{N(N+mn)}, \quad i,j \right) \\ - \sum_{i=1}^m r_i \ln r_i - \sum_{j=1}^n c_j \ln c_j \\ + N \ln N + (N+mn) \ln (N+mn) \end{aligned} \tag{2.3.3}$$

Let us consider the $m \times n$ matrix with the (i,j)-th entry equal to $(r_i c_j + N)/(N^2+Nmn)$. The i-th row sum of the matrix is $(r_i+n)/(N+mn)$, the j-th column sum is $(c_j+m)/(N+mn)$ while the sum of all the entries of the matrix is 1. Using the inequality relating the entropies of two partitions of a probability space with the entropy of the intersection of the partition (see, for example, [Khi57]), we conclude that

$$\begin{aligned} &\mathbf{H} \left(\frac{r_i c_j + N}{N(N+mn)}, \quad \begin{matrix} 1 \leq i \leq m \\ 1 \leq j \leq n \end{matrix} \right) \\ &\leq \mathbf{H} \left(\frac{r_i+n}{N+mn}, \ 1 \leq i \leq m \right) \\ &\quad + \mathbf{H} \left(\frac{c_j+m}{N+mn}, \ 1 \leq j \leq n \right) \end{aligned} \tag{2.3.4}$$

with the equality if and only if

$$\frac{r_i+n}{N+mn} \cdot \frac{c_j+m}{N+mn} = \frac{r_i c_j + N}{N(N+mn)} \quad \text{for all} \quad i,j. \tag{2.3.5}$$

Identities (2.3.5) are equivalent to $(N - r_i m)(N - c_j n) = 0$, which, in turn, equivalent to all row sums being equal $r_i = N/m$ or all column sums being equal $c_j = N/n$.

Summarizing (2.2.1), (2.2.2), (2.3.2), and (2.3.3) we conclude that inequality (2.3.1) indeed holds if not all row sums r_i are equal and not all column sums c_j are equal. Therefore, in the space of $km \times kn$ matrices with the sum $k^2 N$ of all entries the two events

$$\begin{aligned} &\mathcal{R}_k : \quad \text{the vector of row sums of a matrix is} \quad R_k \\ &\qquad\qquad \text{and} \\ &\mathcal{C}_k : \quad \text{the vector of column sums of a matrix is} \quad C_k, \end{aligned} \tag{2.3.6}$$

instead of being asymptotically independent, attract exponentially in k^2, that is,

$$\frac{\mathbf{Pr}\ (\mathcal{R}_k \cap \mathcal{C}_k)}{(\mathbf{Pr}\,\mathcal{R}_k)\,(\mathbf{Pr}\,\mathcal{C}_k)} \ \geq\ \gamma^{k^2}$$

for some $\gamma = \gamma(R, C) > 1$ and all sufficiently large k.

Starting with non-constant margins (R, C) the cloning procedure $(R, C) \longmapsto (R_k, C_k)$ produces margins which stay away from from constant and maintain the density N/mn separated from 0. Similar analysis shows that the phenomenon of attraction of the events $\mathcal{R}_k$ and $\mathcal{C}_k$ defined by (2.3.6) holds for more general sequences of margins (R_k, C_k) of growing dimensions which stay sufficiently away from uniform and sparse.

Two terms contribute to the difference $\ln \#(R, C) - \ln I(R, C)$:

first, the difference $g(Z) - g(Y)$, where Z is the matrix of Lemma 1.4 at which the maximum of the function $g(X) = \sum_{ij} (x_{ij} + 1) \ln(x_{ij} + 1) - x_{ij} \ln x_{ij}$ on the transportation polytope $\mathcal{P}(R, C)$ is attained and $Y = (y_{ij})$ is the independence matrix $y_{ij} = r_i c_j / N$, cf. (2.3.2);

and second, the difference (multiplied by $(N + mn)$) between the entropies on the right hand side of (2.3.4) and the left hand side of (2.3.4).

As long as either of these differences remains large enough to overcome the error term of $O\big((m+n) \ln N\big)$ coming from Theorem 1.1, we have the asymptotic positive correlation of sequences of events $\mathcal{R}_k$ and $\mathcal{C}_k$ in (2.3.6).

On the other hand, the independence estimate $I(R, C)$ produces a reasonable approximation to $\#(R, C)$ in the cases of sparse tables (cf. [GM07]) and tables with constant margins (cf. [CM07a]). One can show that if all row sums are equal or if all column sums are equal then indeed

$$\lim_{k \longrightarrow +\infty} \frac{1}{k^2} \ln \#(R_k, C_k) \ =\ \lim_{k \longrightarrow +\infty} \frac{1}{k^2} \ln I(R_k, C_k),$$

where (R_k, C_k) are cloned margins (R, C). Indeed, if all r_i are equal then the symmetry argument shows that the matrix $Z = (z_{ij})$ in Lemma 1.4 satisfies $z_{ij} = c_j/m$ for all i and j, and, similarly, if all c_j are equal then we have $z_{ij} = r_i/n$ for all i, j. In either case we have $Z = Y$ in (2.3.2) and, as we have already discussed, equations (2.3.5) hold as well.

3. The volume of a section of a simplex

Let $\mathcal{A}$ be the affine hyperplane in $\mathbb{R}^d$ defined by the equation

$$\sum_{i=1}^{d} x_i = 1$$

and let $\Delta \subset \mathcal{A}$ be the standard $(d-1)$-dimensional open simplex defined by the inequalities

$$x_i > 0 \quad \text{for} \quad i = 1, \ldots, d.$$

We consider the Euclidean structure in $\mathcal{A}$ induced from $\mathbb{R}^d$. In particular, if $K \subset \mathcal{A}$ is an m-dimensional convex body, by $\mathrm{vol}_m(K)$ we denote the m-dimensional volume of K with respect to that Euclidean structure. For $m = d-1$ we denote vol_m just by vol. In particular,

$$\mathrm{vol}(\Delta) = \frac{\sqrt{d}}{(d-1)!}.$$

Let $L \subset \mathcal{A}$ be an affine subspace intersecting Δ. Suppose that $\dim L = d-k-1$, so the the codimension of L in $\mathcal{A}$ is $k \geq 1$. Our aim is to estimate the volume of the intersection $\mathrm{vol}_{d-k-1}(L \cap \Delta)$ within a reasonable accuracy when the codimension k of L is small. It turns out that the volume is controlled by one particular quantity, namely the maximum value of the product of the coordinates of a point $x \in \Delta \cap L$.

Our result is as follows.

(3.1) Theorem. *Let $\mathcal{A} \subset \mathbb{R}^d$ be the affine hyperplane defined by the equation $x_1 + \ldots + x_d = 1$ and let $\Delta \subset \mathcal{A}$ be the standard $(d-1)$-dimensional open simplex defined by the inequalities $x_1 > 0, \ldots, x_d > 0$.*

Let $L \subset \mathcal{A}$ be an affine subspace intersecting Δ and such that $\dim L = d-k-1$ where $k \geq 1$. Suppose that the maximum of the function

$$f(x) = \sum_{i=1}^{d} \ln x_i$$

on $\Delta \cap L$ is attained at $a = (\alpha_1, \ldots, \alpha_d)$.

(1) *We have*

$$\frac{\mathrm{vol}_{d-k-1}(\Delta \cap L)}{\mathrm{vol}(\Delta)} \geq \gamma \frac{1}{d^2 \omega_k} d^d e^{f(a)},$$

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 (one can choose $\gamma = 1/2e^3 \approx 0.025$).

(2) *Suppose that*

$$\alpha_i \geq \frac{\epsilon}{d} \quad \textit{for some} \quad \epsilon > 0 \quad \textit{and} \quad i = 1, \ldots, d.$$

Then

$$\frac{\operatorname{vol}_{d-k-1}(\Delta \cap L)}{\operatorname{vol}(\Delta)} \ \leq \ \gamma \left(\frac{d^2}{2\epsilon}\right)^k d^d e^{f(a)},$$

where $\gamma > 0$ is an absolute constant (one can choose $\gamma = 2e \approx 5.44$).

We are interested in the situation of $k \sim \sqrt{d}$, so ignoring lower-order terms in the logarithmic order, we get

$$\operatorname{vol}_{d-k-1}(\Delta \cap L) \sim \frac{d^d}{d!} e^{f(a)} \sim e^d \prod_{i=1}^d \alpha_i,$$

provided the maximum value of the product of the coordinates of a point $x \in \Delta \cap L$ is attained at $a = (\alpha_1, \ldots, \alpha_d)$ and all α_i are not too small.

Let

$$c = \left(\frac{1}{d}, \ldots, \frac{1}{d}\right)$$

be the center of the simplex Δ.

We deduce Theorem 3.1 from the following result.

(3.2) Theorem. *Let $\mathcal{A} \subset \mathbb{R}^d$ be the affine hyperplane defined by the equation $x_1 + \ldots + x_d = 1$ and let $\Delta \subset \mathcal{A}$ be the standard $(d-1)$-dimensional open simplex defined by the inequalities $x_1 > 0, \ldots, x_d > 0$.*

Let $H \subset \mathcal{A}$ be an affine hyperplane in $\mathcal{A}$ intersecting Δ. If H does not pass through the center c of Δ, let $H^- \subset \mathcal{A}$ be the open halfspace bounded by H that does not contain c and if H passes through c let $H^- \subset \mathcal{A}$ be either of the open halfspaces bounded by H.

Suppose that the function

$$f(x) = \sum_{i=1}^d \ln x_i \quad \textit{where} \quad x = (x_1, \ldots, x_d)$$

attains its maximum on $\Delta \cap H$ at a point $a = (\alpha_1, \ldots, \alpha_d)$.

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

$$d^d e^{f(a)} \geq \frac{\operatorname{vol}(\Delta \cap H^-)}{\operatorname{vol}(\Delta)} \geq \frac{\gamma}{d^2} d^d e^{f(a)}.$$

We can choose $\gamma = 1/2e^3 \approx 0.025$.

4. Preliminaries from convex geometry

We recall that $\mathcal{A} \subset \mathbb{R}^d$ is the affine hyperplane defined by the equation $x_1 + \ldots + x_d = 1$, that $\Delta \subset \mathcal{A}$ is the open simplex defined by the inequalities $x_i > 0$ for $i = 1, \ldots, d$, and that $c = (1/d, \ldots, 1/d)$ is the center of Δ. We need some results regarding central hyperplane sections of Δ.

(4.1) Lemma. *Let $H \subset \mathcal{A}$ be an affine hyperplane in $\mathcal{A}$ passing through the center c of Δ.*

(1) *Let H^+ and H^- be the open halfspaces bounded by H. Then for some absolute constant $\gamma > 0$ we have*

$$\frac{\operatorname{vol}(\Delta \cap H^+)}{\operatorname{vol}(\Delta)} \geq \gamma \quad \textit{and} \quad \frac{\operatorname{vol}(\Delta \cap H^-)}{\operatorname{vol}(\Delta)} \geq \gamma.$$

One can choose $\gamma = 1/e \approx 0.37$.

(2) *For some absolute constant $\gamma > 0$ we have*

$$\frac{\operatorname{vol}_{d-2}(\Delta \cap H)}{\operatorname{vol}(\Delta)} \geq \gamma.$$

One can choose $\gamma = 1/2e \approx 0.18$.

Proof. Part (1) is a particular case of a more general result of Grünbaum [Grü60] on hyperplane sections through the centroid of a convex body. In fact, in dimension d one can choose

$$\gamma_d = \left(1 - \frac{1}{d}\right)^{d-1} > \frac{1}{e}.$$

As K. Ball and M. Fradelizi explained to the author, a stronger estimate than that of Part (2) can be obtained by combining techniques of [Bal88] and [Frad97]. Nevertheless, we present a proof of Part (2) below since the same approach is used later in the proof of Theorem 3.1.

To prove Part (2), let $H^\perp \subset \mathcal{A}$ be a line orthogonal to H. Let us consider the orthogonal projection $pr : \mathcal{A} \longrightarrow H^\perp$ and let $Q = pr(\Delta)$ be the image of the simplex. Since Δ is contained in a ball of radius 1, Q is an interval of length at most 2.

Let $y_0 = pr(H)$ and for $y \in Q$ let

$$\nu(y) = \operatorname{vol}_{d-2}\left(pr^{-1}(y)\right)$$

be the volume of the inverse image of y. By the Brunn-Minkowski inequality, the function ν is log-concave, see [Bal88], [Bal97].

Our goal is to bound $\nu(y_0)$ from below. The point y_0 splits the interval Q into two subintervals, $Q^+ = pr(\Delta \cap H^+)$ and $Q^- = pr(\Delta \cap H^-)$ of length at most 2 each.

We have

$$\int_{Q^+} \nu(y)\ dy = \operatorname{vol}\left(\Delta \cap H^+\right) \quad \text{and} \quad \int_{Q^-} \nu(y)\ dy = \operatorname{vol}\left(\Delta \cap H^-\right).$$

Using Part (1) we conclude that there exist $y^+ \in Q^+$ and $y^- \in Q^-$ such that

$$\frac{\nu(y^+)}{\operatorname{vol}(\Delta)} \geq \frac{1}{2e} \quad \text{and} \quad \frac{\nu(y^-)}{\operatorname{vol}(\Delta)} \geq \frac{1}{2e}.$$

Since y_0 is a convex combination of y^+ and y^-, by the log-concavity of ν we must have

$$\frac{\nu(y_0)}{\operatorname{vol}(\Delta)} \geq \frac{1}{2e},$$

as desired. □

Let us choose a point $a = (\alpha_1, \ldots, \alpha_d)$ in Δ, and let us consider the *projective transformation* $T_a : \Delta \longrightarrow \Delta$

$$T_a(x) = y \quad \text{where} \quad y_i = \frac{\alpha_i x_i}{\alpha_1 x_1 + \ldots + \alpha_d x_d} \quad \text{for}$$
$$x = (x_1, \ldots, x_d) \quad \text{and} \quad y = (y_1, \ldots, y_d).$$

The inverse transformation is T_b for $b = \left(\alpha_1^{-1}, \ldots, \alpha_d^{-1}\right)$. Clearly,

$$T_a(c) = a,$$

where c is the center of Δ. For $x \in \Delta$, the derivative $DT_a(x)$ is a linear transformation

$$DT_a(x) : \quad \mathcal{H} \longrightarrow \mathcal{H},$$

where $\mathcal{H}$ is the hyperplane $x_1 + \ldots + x_d = 0$ in $\mathbb{R}^d$.

Our immediate goal is to compute the Jacobian $|DT_a(x)|$ at $x \in \Delta$.

(4.2) Lemma. *Let us choose a point $a = (\alpha_1, \ldots, \alpha_d)$ in the simplex Δ and let us consider the projective transformation*

$$T_a : \Delta \longrightarrow \Delta$$

defined by the formula

$$T_a(x) = y \quad \textit{where} \quad y_i = \frac{\alpha_i x_i}{\alpha_1 x_1 + \ldots + \alpha_d x_d}$$

for $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$.

Let $DT_a(x) : \mathcal{H} \longrightarrow \mathcal{H}$ be the derivative of T_a at $x \in \Delta$ and $|DT_a(x)|$ the corresponding value of the Jacobian. Then

$$|DT_a(x)| = \frac{\alpha_1 \cdots \alpha_d}{(\alpha_1 x_1 + \ldots + \alpha_d x_d)^d}.$$

Proof. Let us consider T_a as defined in a neighborhood of x in $\mathbb{R}^d$ with values in $\mathbb{R}^d$ and let $DT_a(x)$ be the $d \times d$ matrix of the derivative

$$DT_a(x) = \left(\frac{\partial y_i}{\partial x_j} \right)$$

in the standard basis of $\mathbb{R}^d$. Then the i-th diagonal entry of $DT_a(x)$ is

$$\frac{\alpha_i}{(\alpha_1 x_1 + \ldots + \alpha_d x_d)} - \frac{\alpha_i^2 x_i}{(\alpha_1 x_1 + \ldots + \alpha_d x_d)^2},$$

while the (i,j)-th entry for $i \neq j$ is

$$\frac{-\alpha_i \alpha_j x_i}{(\alpha_1 x_1 + \ldots + \alpha_d x_d)^2}.$$

Let

$$\beta = \frac{1}{\alpha_1 x_1 + \ldots + \alpha_d x_d},$$

let B be the diagonal matrix with the diagonal entries $\alpha_1, \ldots, \alpha_d$, and let C be the matrix with the (i,j)th entry equal to $\alpha_i \alpha_j x_i$ for all $1 \leq i, j \leq d$. Then we can write

$$DT_a(x) = \beta B - \beta^2 C = \beta(B - \beta C).$$

Since $DT_a(x)$ maps $\mathbb{R}^d$ onto $\mathcal{H}$ and $\mathcal{H}$ is an invariant subspace of $DT_a(x)$, the value of the Jacobian we are interested in is the product of the $(d-1)$ non-zero eigenvalues of $DT_a(x)$ (counting algebraic multiplicities), which is equal to the $(d-1)$-st elementary symmetric function in the eigenvalues of $DT_a(x)$, which is equal to the sum of the d of the $(d-1) \times (d-1)$ principle minors of $DT_a(x)$.

Let B_i and C_i be the $(d-1) \times (d-1)$ matrices obtained from B and C respectively by crossing out the ith row and column.

Hence

$$B_i - \beta C_i = B_i (I - \beta B_i^{-1} C_i),$$

where I is the $(d-1) \times (d-1)$ identity matrix. Now $B_i^{-1} C_i$ is a matrix of rank 1 with the non-zero eigenvalue equal to the trace of $B_i^{-1} C_i$, which is

$$\sum_{j \neq i} \alpha_j x_j.$$

Hence

$$\det\left(I-\beta B_i^{-1}C_i\right)=1-\beta\sum_{j\neq i}\alpha_j x_j=\frac{\alpha_i x_i}{\alpha_1 x_1+\ldots+\alpha_d x_d}$$

and

$$\det B_i(I-\beta B_i^{-1}C_i)=\frac{\alpha_1\cdots\alpha_d x_i}{\alpha_1 x_1+\ldots+\alpha_d x_d}.$$

Therefore, the sum of the d of $(d-1)\times(d-1)$ principle minors of $B-\beta C$ is

$$\frac{\alpha_1\ldots\alpha_d}{\alpha_1 x_1+\ldots+\alpha_d x_d}$$

and the sum of the $(d-1)\times(d-1)$ principle minors of $DT_a(x)=\beta(B-\beta C)$ is

$$\frac{\alpha_1\ldots\alpha_d}{(\alpha_1 x_1+\ldots+\alpha_d x_d)^d},$$

as desired. □

Next, we will need a technical estimate, which shows that if the volume of the section of the simplex by an affine subspace of a small codimension is sufficiently large and if the subspace cuts sufficiently deep into the simplex then a neighborhood of the section in the simplex has a sufficiently large volume.

(4.3) Lemma. *Let $L\subset\mathcal{A}$ be an affine subspace,* $\dim L=d-k-1$. *Suppose that there is a point $a\in L\cap\Delta$, $a=(\alpha_1,\ldots,\alpha_d)$ such that*

$$\alpha_i\geq\frac{\epsilon}{d}\quad\text{for}\quad i=1,\ldots,d$$

and some $\epsilon>0$.

Let

$$\|x\|_\infty=\max\bigl\{|x_i|\quad\text{for}\quad i=1,\ldots,d\bigr\}\quad\text{for}\quad i=1,\ldots,d$$

and let us define a neighborhood Q of $\Delta\cap L$ by

$$Q=\Bigl\{x\in\Delta:\quad\|x-y\|_\infty\leq\frac{\epsilon}{d^2}\quad\text{for some}\quad y\in\Delta\cap L\Bigr\}.$$

Then, for any affine hyperplane $H\subset\mathcal{A}$ passing through L we have

$$\operatorname{vol}(Q\cap H^+),\quad\operatorname{vol}(Q\cap H^-)\geq\gamma\left(\frac{2\epsilon}{d^2}\right)^k\operatorname{vol}_{d-k-1}(\Delta\cap L),$$

where H^+ and H^- are the halfspaces bounded by H and $\gamma>0$ is an absolute constant. One can choose $\gamma=1/2e\approx 0.18$.

Proof. Let

$$Q_0=\left\{\frac{1}{d}a+\frac{d-1}{d}x:\quad x\in\Delta\cap L\right\}.$$

Since Q_0 is the contraction of $\Delta \cap L$ we have

$$\mathrm{vol}_{d-k-1} Q_0 = \left(\frac{d-1}{d}\right)^{d-k-1} \mathrm{vol}_{d-k-1}(\Delta \cap L) \geq \frac{1}{e} \mathrm{vol}_{d-k-1}(\Delta \cap L).$$

Moreover, for any $x \in Q_0$, $x = (x_1, \ldots, x_d)$, we have

$$x_i \geq \frac{\epsilon}{d^2} \quad \text{for} \quad i = 1, \ldots, d.$$

For every point $x \in Q_0$ let us consider the cube

$$I_x = \left\{ y \in \mathbb{R}^d : \quad \|y - x\|_\infty \leq \frac{\epsilon}{d^2} \right\}.$$

Then $(I_x \cap \mathcal{A}) \subset \Delta$. The intersection of I_x with the k-dimensional affine subspace $L_x^\perp \subset \mathcal{A}$ orthogonal to L and passing through x is centrally symmetric with respect to x and, by Vaaler's Theorem [Vaa79], satisfies

$$\mathrm{vol}_k \left(I_x \cap L_x^\perp\right) \geq \left(\frac{2\epsilon}{d^2}\right)^k.$$

The proof now follows. □

5. Proofs of Theorems 1.2, 3.1, and 3.2

We prove Theorem 3.2 first.

(5.1) Proof of Theorem 3.2. If $c \in H$ the result follows by Lemma 4.1. Hence we assume that $c \notin H$.

The hyperplane H is orthogonal to the gradient of $f(x)$ at $x = a$ and passes through a, from which it follows that H can be defined in $\mathcal{A}$ by the equation

$$\sum_{i=1}^d \frac{x_i}{\alpha_i} = d,$$

while the halfspace H^- is defined by the inequality

$$\sum_{i=1}^d \frac{x_i}{\alpha_i} < d.$$

Let us consider the projective transformation $T_a : \Delta \longrightarrow \Delta$ defined by the formula of Lemma 4.2. Hence $T_a(c) = a$. Moreover, the inverse image $T_a^{-1}(H)$ is the hyperplane H_0 defined in $\mathcal{A}$ by the equation

$$\sum_{i=1}^d \alpha_i x_i = \frac{1}{d}$$

and the inverse image $T_a^{-1}(\Delta \cap H^-)$ is the intersection $\Delta \cap H_0^-$, where H_0^- is the halfspace defined by the inequality

$$\sum_{i=1}^d \alpha_i x_i > \frac{1}{d}.$$

By Lemma 4.2, we have

$$|DT_a(x)| = \frac{\alpha_1 \cdots \alpha_d}{(\alpha_1 x_1 + \ldots + \alpha_d x_d)^d} < d^d \prod_{i=1}^d \alpha_i \quad \text{for all} \quad x \in \Delta \cap H_0^-.$$

Since

$$\text{vol}(\Delta \cap H^-) = \int_{\Delta \cap H_0^-} |DT_a(x)| \, dx, \tag{5.1.1}$$

the upper bound follows.

Let us prove the lower bound. By Part (2) of Lemma 4.1,

$$\frac{\text{vol}_{d-2}(\Delta \cap H_0)}{\text{vol}(\Delta)} \geq \frac{1}{2e}.$$

We recall that H_0 passes through the center of the simplex and apply Lemma 4.3 with $\epsilon = 1$. Namely, we define

$$Q = \left\{ x \in \Delta : \quad \|x - y\|_\infty \leq \frac{1}{d^2} \quad \text{for some} \quad y \in \Delta \cap H_0 \right\}$$

and conclude that by Lemma 4.3

$$\text{vol}(Q \cap H_0^-) \;\geq\; \left(\frac{1}{2e}\right)\left(\frac{2}{d^2}\right) \text{vol}_{d-2}(\Delta \cap H_0) \;\geq\; \frac{1}{2e^2 d^2} \text{vol}(\Delta).$$

We note that for every $x \in Q$ we have

$$\sum_{i=1}^d \alpha_i x_i \leq \frac{1}{d} + \frac{1}{d^2} = \frac{d+1}{d^2}.$$

By (5.1.1)

$$\begin{aligned} \text{vol}(\Delta \cap H^-) \geq \int_{Q \cap H_0^-} |DT_a(x)| \, dx \geq \left(\frac{d^2}{d+1}\right)^d \text{vol}(Q \cap H_0^-) \prod_{i=1}^d \alpha_i \\ \geq \frac{1}{2e^3} \frac{1}{d^2} d^d \, \text{vol}(\Delta) \prod_{i=1}^d \alpha_i, \end{aligned}$$

which completes the proof. □

Next, we prove Theorem 3.1.

(5.2) Proof of Theorem 3.1. The proof of Part (1) is similar to that of Part (2) of Lemma 4.1. Let $L^{\perp} \subset \mathcal{A}$ be a k-dimensional subspace orthogonal to L in $\mathcal{A}$ and let

$$pr : \mathcal{A} \longrightarrow L^{\perp}$$

be the orthogonal projection. Let $Q \subset L^{\perp}$, $Q = pr(\Delta)$, be the image of the simplex. Clearly, Q lies in a ball of radius 1, so

$$\operatorname{vol}_k Q \ \leq \ \omega_k.$$

For $y \in Q$, let

$$\nu(y) = \operatorname{vol}_{d-k-1}\left(pr^{-1}(y)\right)$$

be the volume of the inverse image of y. By the Brunn-Minkowski inequality, the function ν is log-concave, so for every $\alpha > 0$ the set

$$\Big\{ y \in Q : \quad \nu(y) \geq \alpha \Big\}$$

is convex. Moreover, for all Borel sets $Y \subset Q$ we have

$$\int_Y \nu(y) \ dy = \operatorname{vol}\left(pr^{-1}(Y)\right).$$

We want to estimate $\nu(y_0)$ for $y_0 = pr(L)$. Let $H \subset L^{\perp}$ be an affine hyperplane through y_0 and let $H^+, H^- \subset L^{\perp}$ be open halfspaces bounded by H. Then $\tilde{H} = pr^{-1}(H)$ is an affine hyperplane in $\mathcal{A}$ containing L and $pr^{-1}(H^-)$ and $pr^{-1}(H^+)$ are the corresponding open halfspaces of $\mathcal{A}$ bounded by $\tilde{H}$.

Since the maximum value of f on $\Delta \cap \tilde{H}$ is at least as big as the maximum value of f on $\Delta \cap L$, by Theorem 3.2 we have

$$\operatorname{vol} pr^{-1}\left(H^{\pm} \cap Q\right) \geq \frac{1}{2e^3} \frac{1}{d^2} d^d e^{f(a)} \operatorname{vol}(\Delta).$$

Since

$$\operatorname{vol} pr^{-1}\left(H^{\pm} \cap Q\right) = \int_{H^{\pm} \cap Q} \nu(y) \ dy,$$

We conclude that there exist points $y^+ \in H^+$ and $y^- \in H^-$ such that

$$\nu(y^+), \ \nu(y^-) \ \geq \ \frac{1}{2e^3} \frac{1}{d^2} d^d e^{f(a)} \frac{\operatorname{vol}(\Delta)}{\operatorname{vol}_k Q} \ \geq \ \frac{1}{2e^3} \frac{1}{d^2} d^d e^{f(a)} \frac{\operatorname{vol}(\Delta)}{\omega_k}. \tag{5.2.1}$$

In other words, for any affine hyperplane $H \subset L^{\perp}$ through y_0 on either side of the hyperplane there are points y^+, y^- for which inequality (5.2.1) holds. Hence y_0 lies in the convex hull of points y for which the inequality holds. The proof of Part (1) follows by the log-concavity of ν.

Let us prove Part (2). Since a is the maximum point of the strictly concave function

$$f(x)=\sum_{i=1}^{d}\ln x_i$$

on $\Delta\cap L$, the gradient of f at a is orthogonal to L. Hence L is orthogonal to the vector

$$\left(\frac{1}{\alpha_1},\dots,\frac{1}{\alpha_d}\right).$$

If $a\neq c$, let $H\subset\mathcal{A}$ be the affine hyperplane defined by the equation

$$\sum_{i=1}^{d}\frac{x_i}{\alpha_i}=d$$

and if $a=c$ let H be any affine hyperplane containing L. In either case $L\subset H$ and the maximum values of f on $\Delta\cap H$ and on $\Delta\cap L$ coincide and are equal to $f(a)$. Therefore, by Theorem 3.2, we have

$$\text{vol}(\Delta\cap H^-)\ \leq\ d^d e^{f(a)}\,\text{vol}(\Delta) \tag{5.2.2}$$

for some open halfspace H^- bounded by H.

We apply Lemma 4.3. Namely, we let

$$Q=\left\{x\in\Delta:\quad \|x-y\|_\infty\leq\frac{\epsilon}{d^2}\quad\text{for some}\quad y\in\Delta\cap L\right\}.$$

Then, by Lemma 4.3,

$$\text{vol}(Q\cap H^-)\ \geq\ \frac{1}{2e}\left(\frac{2\epsilon}{d^2}\right)^k \text{vol}_{d-k-1}(\Delta\cap L).$$

Since

$$\text{vol}(Q\cap H^-)\ \leq\ \text{vol}(\Delta\cap H^-),$$

we get the upper bound from (5.2.2). □

Finally, we prove Theorem 1.2.

(5.3) Proof of Theorem 1.2. Let us consider the contracted polytope $N^{-1}\mathcal{P}$ defined by the equations

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

and inequalities

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

Then $N^{-1}\mathcal{P}$ can be represented as an intersection of the standard simplex in the space of $m \times n$ matrices and an affine subspace of dimension $(m-1)(n-1)$. We are going to use Theorem 3.1. Let $A = (\alpha_{ij})$, $A \in N^{-1}\mathcal{P}$, be the point maximizing the product of the coordinates. Writing the optimality condition for

$$f(X) = \sum_{ij} \ln x_{ij}$$

on $N^{-1}\mathcal{P}$, we conclude that

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

and some $\lambda_1, \ldots \lambda_m$ and $\mu_1, \ldots, \mu_n$. Since $\lambda_i + \mu_j > 0$ for all i, j, we may assume that $\lambda_i, \mu_j > 0$ for all i, j. If $\lambda_i > nN/r_i$ for some i then $\alpha_{ij} < r_i/nN$ for all j, which is a contradiction. If $\mu_j > mN/c_j$ for some j then $\alpha_{ij} < c_j/mN$ for all i which is a contradiction. Hence $\lambda_i \leq nN/r_i$ for $i = 1, \ldots, m$ and $\mu_j \leq mN/c_j$ for $j = 1, \ldots, n$, from which

$$\alpha_{ij} \geq \frac{r_i c_j}{Nnc_j + Nmr_i} = \frac{1}{(nN/r_i) + (mN/c_j)} \quad \text{for all} \quad i, j.$$

The proof now follows by Theorem 3.1 with $d = mn$, $k = m + n - 2$, and

$$\epsilon = \left(n \max_{i=1,\ldots,m} \frac{N}{r_i} + m \max_{j=1,\ldots,n} \frac{N}{c_j} \right)^{-1}.$$

□

6. An integral representation for the number of contingency tables

In this section, we recall bounds for $\#(R, C)$ obtained in [Ba07] and [Ba08].

(6.1) Matrix scaling. Our estimates for the number $\#(R, C)$ of contingency tables essentially use the theory of *matrix scaling*, see [Si64], [MO68], [RS89]. Let us fix non-negative vectors $R = (r_1, \ldots, r_m)$, $C = (c_1, \ldots, c_n)$, such that

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

Then for every $m \times n$ positive matrix $X = (x_{ij})$ there exist a positive $m \times n$ matrix $L = (l_{ij})$ and positive numbers $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$ such that

$$\begin{aligned} &\sum_{j=1}^n l_{ij} = r_i \quad \text{for} \quad i = 1, \ldots, m, \\ &\sum_{i=1}^m l_{ij} = c_j \quad \text{for} \quad j = 1, \ldots, n, \qquad \text{and} \\ &x_{ij} = \lambda_i \mu_j l_{ij} \quad \text{for all} \quad i, j. \end{aligned} \tag{6.1.1}$$

Moreover, given X, the matrix L is unique while the numbers λ_i and μ_j are unique up to a re-scaling:

$$\lambda_i \longmapsto \lambda_i \tau \quad \text{for} \quad i=1, \ldots, m$$
$$\mu_j \longmapsto \mu_j \tau^{-1} \quad \text{for} \quad j=1, \ldots, n$$

and some $\tau > 0$.

(6.2) Function ϕ. This allows us to define a function

$$\phi(X) = \phi_{R,C}(X) = \left(\prod_{i=1}^{m} \lambda_i^{r_i} \right) \left(\prod_{j=1}^{n} \mu_j^{c_j} \right),$$

where λ_i and μ_j are numbers such that equations (6.1.1) hold, on positive $m \times n$ matrices X. It turns out that ϕ is continuous (it is also log-concave but we don't use that), positive homogeneous of degree N,

$$\phi(\alpha X) = \alpha^N \phi(X)$$

for $\alpha > 0$ and positive matrix X, and monotone

$$\phi(X) \geq \phi(Y)$$

provided X and Y are positive matrices satisfying $x_{ij} \geq y_{ij}$ for all i, j, see, for example, [Ba07] and [Ba08].

Alternatively, $\phi(X)$ can be defined by

$$\phi(X) = \min_{a,b} \left(\frac{1}{N} \sum_{ij} x_{ij} \alpha_i \beta_j \right)^N,$$

where the minimum is taken over all positive m-vectors $a = (\alpha_1, \ldots, \alpha_m)$ and positive n-vectors $b = (\beta_1, \ldots, \beta_n)$ satisfying

$$\prod_{i=1}^{m} \alpha_i^{r_i} = \prod_{j=1}^{n} \beta_j^{c_j} = 1,$$

see also [MO68].

(6.3) The bounds. Let us identify the space of $m \times n$ matrices with Euclidean space $\mathbb{R}^d$ for $d = mn$, let $\mathcal{A} \subset \mathbb{R}^d$ be the affine hyperplane defined by the equation

$$\sum_{ij} x_{ij} = 1,$$

and let $\Delta \subset \mathcal{A}$ be the standard open simplex defined by the inequalities

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

with the Lebesgue measure dX induced from the Euclidean structure in $\mathbb{R}^d$. It is proved in [Ba07] and [Ba08] that

$$\begin{gathered}
\#(R, C) \ \geq \ \frac{N!(N+mn-1)!}{N^N \sqrt{mn}} \left(\prod_{i=1}^m \frac{r_i^{r_i}}{r_i!} \right) \left(\prod_{j=1}^n \frac{c_j^{c_j}}{c_j!} \right) \int_{\Delta} \phi(X) \ dX \\
\text{and} \\
\#(R, C) \ \leq \ \frac{(N+mn-1)!}{\sqrt{mn}} \min \left\{ \prod_{i=1}^m \frac{r_i^{r_i}}{r_i!}, \quad \prod_{j=1}^n \frac{c_j^{c_j}}{c_j!} \right\} \int_{\Delta} \phi(X) \ dX.
\end{gathered} \tag{6.3.1}$$

Therefore, we have an approximation within up to an $N^{\gamma(m+n)}$ factor for some absolute constant $\gamma > 0$:

$$\#(R, C) \approx e^N (N+mn)! \int_{\Delta} \phi(X) \ dX \tag{6.3.2}$$

In fact, we will be using only a lower bound in (6.3.1).

For completeness, let us sketch the main ingredients of the proof of (6.3.1).

Recall that the *permanent* of an $N \times N$ matrix $A = (a_{ij})$ is defined by the formula

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

where the sum is taken over all $N!$ permutations σ from the symmetric group S_N. For an $m \times n$ matrix $X = (x_{ij})$ let us define the $N \times N$ block matrix $A(X)$ that has mn blocks of sizes $r_i \times c_j$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$ with the (i, j)-th block filled by the copies of x_{ij}. A combinatorial computation produces the following expansion

$$\operatorname{per} A(X) = \left(\prod_{i=1}^m r_i! \right) \left(\prod_{j=1}^n c_j! \right) \sum_{D = (d_{ij})} \frac{x_{ij}^{d_{ij}}}{d_{ij}!},$$

where the sum is taken over all $m \times n$ non-negative integer matrices $D = (d_{ij})$ with row sums R and column sums C. From this expansion we obtain the formula

$$\#(R, C) = \left(\prod_{i=1}^m \frac{1}{r_i!} \right) \left(\prod_{j=1}^n \frac{1}{c_j!} \right) \int_{\mathbb{R}_+^d} \operatorname{per} A(X) \exp \left\{ -\sum_{ij} x_{ij} \right\} \ dX,$$

where $\mathbb{R}^d_+$ is the set of $m \times n$ positive matrices X, see Theorem 1.1 of [Ba08]. Since $\operatorname{per} A(X)$ is a homogeneous polynomial of degree N in X, a standard change of variables results in the formula

$$\#(R, C) = \frac{(N+mn-1)!}{\sqrt{mn}} \left(\prod_{i=1}^{m} \frac{1}{r_i!} \right) \left(\prod_{j=1}^{n} \frac{1}{c_j!} \right) \int_{\Delta} \operatorname{per} A(X) \ dX,$$

cf. Lemma 4.1 of [Ba08]. Given a matrix $X \in \Delta$, let $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$ be its scaling factors so that (6.1.1) holds. Let $B(X)$ be the matrix obtained by dividing the entries in the (i,j)-th block of $A(X)$ by $\lambda_i r_i \mu_j c_j$, so the entries in the (i,j)-th block of $B(X)$ are equal to $l_{ij}/r_i c_j$. Hence

$$\operatorname{per} A(X) = \left(\prod_{i=1}^{m} r_i^{r_i} \right) \left(\prod_{j=1}^{n} c_j^{c_j} \right) \phi(X) \operatorname{per} B(X),$$

cf. Section 3.1 of [Ba08]. Now we notice that $B(X)$ is a *doubly stochastic matrix*, that is, a non-negative matrix with row and column sums equal to 1. The classical estimate for permanents of doubly stochastic matrices conjectured by van der Waerden and proved by Falikman and Egorychev (see [Fa81], [Eg81], and Chapter 12 of [LW01]) asserts that

$$\operatorname{per} B(X) \geq \frac{N!}{N^N}$$

and hence the lower bound in (6.3.1) follows. The upper bound in (6.3.1) follows from the inequality for permanents conjectured by Minc and proven by Bregman, (see [Br73] and Chapter 11 of [LW01]), which results in

$$\operatorname{per} B(X) \ \leq \ \min \left\{ \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}, \quad \prod_{j=1}^{n} \frac{c_j!}{c_j^{c_j}} \right\},$$

since the entries in the (i,j)-th block of $B(X)$ do not exceed $\min\{1/r_i, \ 1/c_j\}$, see Section 5 of [Ba08] for details.

(6.4) Slicing the simplex. The crucial observation which makes the integral

$$\int_{\Delta} \phi(X) \ dX$$

amenable to analysis is that the simplex Δ can be sliced by affine subspaces of codimension $m+n-1$ into sections on which function ϕ remains constant.

Let us choose some positive $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$ and let us consider the affine subspace $L \subset \mathbb{R}^d$ of $m \times n$ matrices $X = (x_{ij})$ satisfying the equations

(6.4.1)
$$\begin{aligned} \sum_{j=1}^{n} \frac{x_{ij}}{\lambda_i \mu_j} &= r_i \quad \text{for} \quad i = 1, \ldots m \\ \sum_{i=1}^{m} \frac{x_{ij}}{\lambda_i \mu_j} &= c_j \quad \text{for} \quad j = 1, \ldots, n \end{aligned}$$

Clearly,

$$\phi(X) = \left(\prod_{i=1}^m \lambda_i^{r_i}\right)\left(\prod_{j=1}^n \mu_j^{c_j}\right) \quad \text{for all} \quad X \in \Delta \cap L. \tag{6.4.2}$$

Moreover, $\dim L = (m-1)(n-1)$.

(6.5) Modification for weighted tables. Similar identities an inequalities hold for weighted tables. For a *positive* matrix $W = (w_{ij})$ of weights, we define the function

$$\phi_{R,C;W}(X) = \phi_{R,C}(Y) \quad \text{where} \quad y_{ij} = w_{ij}x_{ij} \quad \text{for all} \quad i,j$$

and $\phi_{R,C}$ is the unweighted function defined in Section 6.2. Then

$$\begin{aligned} T(R,C;W) \ &\geq \ \frac{N!(N+mn-1)!}{N^N\sqrt{mn}} \\ &\times \left(\prod_{i=1}^m \frac{r_i^{r_i}}{r_i!}\right)\left(\prod_{j=1}^n \frac{c_j^{c_j}}{c_j!}\right)\int_\Delta \phi_{R,C;W}(X)\ dX \\ \text{and} & \\ T(R,C:W) \ &\leq \ \frac{(N+mn-1)!}{\sqrt{mn}} \\ &\times \min\left\{\prod_{i=1}^m \frac{r_i^{r_i}}{r_i!}, \quad \prod_{j=1}^n \frac{c_j^{c_j}}{c_j!}\right\}\int_\Delta \phi_{R,C;W}(X)\ dX, \end{aligned} \tag{6.5.1}$$

see [Ba07], [Ba08], and the proof sketch in Section 6.3.

Let us choose some positive $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$ and let us consider the subspace $L \subset \mathbb{R}^d$ of $m \times n$ matrices $X = (x_{ij})$ satisfying the equations

$$\begin{aligned} \sum_{j=1}^n \frac{w_{ij}x_{ij}}{\lambda_i\mu_j} &= r_i \quad \text{for} \quad i = 1, \ldots m \\ \sum_{i=1}^m \frac{w_{ij}x_{ij}}{\lambda_i\mu_j} &= c_j \quad \text{for} \quad j = 1, \ldots, n \end{aligned} \tag{6.5.2}$$

Clearly,

$$\phi_{R,C;W}(X) = \left(\prod_{i=1}^m \lambda_i^{r_i}\right)\left(\prod_{j=1}^n \mu_j^{c_j}\right) \quad \text{for all} \quad X \in \Delta \cap L. \tag{6.5.3}$$

Moreover, $\dim L = (m-1)(n-1)$.

7. Proofs of Theorems 1.1 and 1.3 and Lemma 1.4

We prove Lemma 1.4 first.

(7.1) Proof of Lemma 1.4. It is straightforward to check that the function

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

is strictly concave for $x>0$. Therefore, the maximum of $g(X;W)$ on $\mathcal{P}(R,C)$ is attained at a single point $Z=(z_{ij})$. Let us show that necessarily $z_{ij}>0$ for all i,j.

Since

$$g'(x;w)=\ln\left(\frac{x+1}{x}\right)+\ln w,$$

the derivative of $g(x;w)$ at $x>0$ is finite and the right derivative at $x=0$ is $+\infty$. Let $Y\in\mathcal{P}(R,C)$ be a matrix with positive entries, for example, $Y=(y_{ij})$ where $y_{ij}=r_ic_j/N$. If $z_{ij}=0$ for some i,j then

$$g\big((1-\epsilon)Z+\epsilon Y;W\big) \ > \ g\big(Z;W\big)$$

for some sufficiently small $\epsilon>0$, which is a contradiction.

Thus $z_{ij}>0$ for all i,j and hence Z lies in the relative interior of $\mathcal{P}(R,C)$. Therefore the gradient of $g(X;W)$ at $X=Z$ is orthogonal to the affine span of $\mathcal{P}(R,C)$, that is,

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

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

Let

$$\xi_i=e^{-\lambda_i}>0 \quad \text{for} \quad i=1,\ldots,m \quad \text{and} \quad \eta_j=e^{-\mu_j}>0 \quad \text{for} \quad j=1,\ldots,n.$$

Then

$$w_{ij}\xi_i\eta_j=\frac{z_{ij}}{z_{ij}+1}<1 \quad \text{for all} \quad i,j,$$

and

(7.1.1)
$$z_{ij}=\frac{w_{ij}\xi_i\eta_j}{1-w_{ij}\xi_i\eta_j} \quad \text{for all} \quad i,j.$$

In particular,

(7.1.2)
$$\begin{aligned}\sum_{j=1}^{n}\frac{w_{ij}\xi_i\eta_j}{1-w_{ij}\xi_i\eta_j}&=r_i \quad \text{for} \quad i=1,\ldots,m \quad \text{and}\\ \sum_{i=1}^{m}\frac{w_{ij}\xi_i\eta_j}{1-w_{ij}\xi_i\eta_j}&=c_j \quad \text{for} \quad j=1,\ldots,n.\end{aligned}$$

Equations (7.1.2) are equivalent to the statement that the point $\mathbf{t}^* = (\lambda_1, \ldots, \lambda_m)$ and $\mathbf{s}^* = (\mu_1, \ldots, \mu_n)$ is a critical point of the function

$$\phi(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^m r_i t_i + \sum_{j=1}^n c_j s_j - \sum_{ij} \ln\left(1 - w_{ij} e^{-t_i - s_j}\right).$$

Since ϕ is convex, the point $(\mathbf{s}^*, \mathbf{t}^*)$ is a minimum point of ϕ and hence the point $\mathbf{x}^* = (\xi_1, \ldots, \xi_m)$ and $\mathbf{y}^* = (\eta_1, \ldots, \eta_n)$ is a point where the infimum of

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

is attained in the region $x_1, \ldots, x_m > 0$, $y_1, \ldots, y_n > 0$, and $w_{ij} x_i y_j < 1$ for all i, j.

Using (7.1.1) and (7.1.2), we conclude that

$$\begin{aligned}
g(Z; W) = &\sum_{ij} (z_{ij} + 1) \ln(z_{ij} + 1) - \sum_{ij} z_{ij} (\ln z_{ij} - \ln w_{ij}) \\
= &- \sum_{ij} \frac{\ln(1 - w_{ij}\xi_i\eta_j)}{1 - w_{ij}\xi_i\eta_j} - \sum_{ij} \frac{w_{ij}\xi_i\eta_j}{1 - w_{ij}\xi_i\eta_j} \ln\left(\frac{\xi_i\eta_j}{1 - w_{ij}\xi_i\eta_j}\right) \\
= &- \sum_{ij} \ln(1 - w_{ij}\xi_i\eta_j) - \sum_{i=1}^m \ln \xi_i \left(\sum_{j=1}^n \frac{w_{ij}\xi_i\eta_j}{1 - w_{ij}\xi_i\eta_j}\right) \\
&- \sum_{j=1}^n \ln \eta_j \left(\sum_{j=1}^n \frac{w_{ij}\xi_i\eta_j}{1 - w_{ij}\xi_i\eta_j}\right) \\
= &- \sum_{i=1}^m r_i \xi_i - \sum_{j=1}^n c_j \eta_j - \sum_{ij} \ln(1 - w_{ij}\xi_i\eta_j) = \ln F(\mathbf{x}^*, \mathbf{y}^*; W),
\end{aligned}$$

as claimed.

We observe that the value of $F(\mathbf{x}, \mathbf{y}; W)$ does not change if we scale $x_i \longmapsto x_i \tau$, $y_j \longmapsto y_j \tau^{-1}$ for $\tau > 0$. In the case of $w_{ij} = 1$ for all i, j we have $\xi_i \eta_j < 1$ for all i, j and hence by choosing an appropriate τ we can enforce $0 < \xi_i, \eta_j < 1$ for all i, j. □

We consider the space $\mathbb{R}^d$ for $d = mn$ of $m \times n$ real matrices, the affine hyperplane $\mathcal{A} \subset \mathbb{R}^d$ defined by the equation $\sum_{ij} x_{ij} = 1$ and the standard open simplex $\Delta \subset \mathcal{A}$ defined by the inequalities $x_{ij} > 0$ for all i, j. Let $\phi = \phi_{R,C;W}$ be the function defined in Sections 6.5 and 6.2.

We start with a technical lemma, which is a straightforward modification of Lemma 4.3.

(7.2) Lemma. *Let $L \subset \mathcal{A}$ be an affine subspace,* $\dim L = d - k - 1$ *for* $k \geq 1$. *Suppose that there is a point $A = (\alpha_{ij})$, $A \in \Delta \cap L$, such that*

$$\alpha_{ij} \geq \frac{\epsilon}{d} \quad \textit{for all} \quad i, j.$$

Suppose further that the value of the function $\phi = \phi_{R,C;W}$ on $\Delta \cap L$ is constant and equal to τ. Then

$$\int_\Delta \phi(X)\ dX \geq \gamma \left(\frac{2\epsilon}{d^2(N+1)} \right)^k \tau \operatorname{vol}_{d-k-1}(\Delta \cap L)$$

for some absolute constant $\gamma > 0$ (one can choose $\gamma = e^{-2} \approx 0.14$).

Proof. Let

$$Q_0 = \left\{ \frac{1}{d} A + \frac{d-1}{d} X : \quad X \in \Delta \cap L \right\}.$$

As in the proof of Lemma 4.3, we have

$$\operatorname{vol}_{d-k-1}(Q_0) \geq \frac{1}{e} \operatorname{vol}_{d-k-1}(\Delta \cap L)$$

and for any $X \in Q_0$, $X = (x_{ij})$, we have

$$x_{ij} \geq \frac{\epsilon}{d^2} \quad \text{for all} \quad i, j.$$

Let us define Q by

$$Q = \left\{ X \in \Delta : \quad \|X - Y\|_\infty \leq \frac{\epsilon}{d^2(N+1)} \quad \text{for some} \quad Y \in Q_0 \right\}.$$

Then, as in Lemma 4.3, we have

$$\operatorname{vol} Q \ \geq \ \frac{1}{e} \left(\frac{2\epsilon}{d^2(N+1)} \right)^k \operatorname{vol}_{d-k-1}(\Delta \cap L).$$

We note that for every $X \in Q$ there is a $Y \in \Delta \cap L$ such that

$$x_{ij} \geq \left(1 - \frac{1}{N+1}\right) y_{ij} \quad \text{for all} \quad i, j.$$

Since ϕ is monotone and homogeneous of degree N (see Section 6.2) , we have

$$\phi(X) \ \geq \ \left(1 - \frac{1}{N+1}\right)^N \tau \ > \ \frac{1}{e}\tau \quad \text{for all} \quad X \in Q.$$

Since

$$\int_\Delta \phi(X)\ dX \geq \int_Q \phi(X)\ dX,$$

the proof follows. □

(7.3) Proof of Theorem 1.1. The upper bound follows immediately from the standard generating function expression:

$$\prod_{i,j} \frac{1}{1-x_i y_j} = \sum_{R,C} \#(R,C) x^R y^C, \quad \text{where} \quad x^R = \prod_{i=1}^m x_i^{r_i} \quad \text{and} \quad x^C = \prod_{j=1}^n y_j^{c_j}$$

and the some is taken over all pairs of positive integer m-vectors $R = (r_1, \ldots, r_m)$ and n-vectors $C = (c_1, \ldots, c_n)$ such that $r_1 + \ldots + r_m = c_1 + \ldots + c_n$.

Let us prove the lower bound. By Lemma 1.4 the minimum of

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

on the open cube $0 < x_i, y_j < 1$ for all i, j is attained at a certain point

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

which, moreover, satisfies

$$\begin{aligned} &\sum_{j=1}^n \frac{\xi_i \eta_j}{1 - \xi_i \eta_j} = r_i \quad \text{for} \quad i = 1, \ldots, m \\ &\sum_{i=1}^m \frac{\xi_i \eta_j}{1 - \xi_i \eta_j} = c_j \quad \text{for} \quad j = 1, \ldots, n. \end{aligned} \tag{7.3.1}$$

Equations (7.3.1) can also be obtained by setting the gradient of $\ln F$ to 0.

In the space of $m \times n$ matrices $\mathbb{R}^d$ with $d = mn$, let us consider the standard simplex Δ and the point $A = (\alpha_{ij})$ defined by

$$\alpha_{ij} = \frac{1}{(N + mn)(1 - \xi_i \eta_j)} \quad \text{for all} \quad i, j.$$

By (7.3.1), we have

$$\sum_{j=1}^n \frac{1}{1 - \xi_i \eta_j} = \sum_{j=1}^n \frac{1 - \xi_i \eta_j}{1 - \xi_i \eta_j} + \sum_{j=1}^n \frac{\xi_i \eta_j}{1 - \xi_i \eta_j} = n + r_i \quad \text{for all} \quad i,$$

so A lies in Δ. Let

$$\begin{aligned} \lambda_i &= \frac{1}{\xi_i \sqrt{N + mn}} \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \\ \mu_j &= \frac{1}{\eta_j \sqrt{N + mn}} \quad \text{for} \quad j = 1, \ldots, n. \end{aligned}$$

Let us consider the affine subspace $L \subset \mathbb{R}^d$ defined by the system of equations

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

Hence $\dim L = (m-1)(n-1)$ and $A \in L$ by (7.3.1).

By (6.4.2), the density $\phi = \phi_{R,C}$ is constant on L and equal to

$$\tau = \frac{1}{(N+mn)^N} \left(\prod_{i=1}^{m} \xi_i^{-r_i} \right) \left(\prod_{j=1}^{n} \eta_j^{-c_j} \right).$$

By Part (1) of Theorem 3.1, the volume of the section $\Delta \cap L$ within a factor of $(N+mn)^{O(m+n)}$ is at least

$$\frac{e^{mn}}{(N+mn)^{mn}} \left(\prod_{ij} \frac{1}{1 - \xi_i \eta_j} \right).$$

More precisely, for $k = d - 1 - \dim(\Delta \cap L)$ we have $k = m+n-1$ or $k = m+n-2$ and

$$\mathrm{vol}_{d-1-k}(\Delta \cap L) \ \geq \ \frac{\Gamma(k/2+1)}{2e^3 \pi^{k/2} \sqrt{mn}} \frac{(mn)^{mn}}{(mn)!} \frac{1}{(N+mn)^{mn}} \left(\prod_{ij} \frac{1}{1 - \xi_i \eta_j} \right).$$

Choosing $\epsilon = 1/(N+mn)$ in Lemma 7.2, we estimate the integral

$$\int_{\Delta} \phi(X) \ dX$$

within a factor of $N^{O(m+n)}$ from below by

$$\frac{e^{mn}}{(N+mn)^{N+mn}} \left(\prod_{i=1}^{m} \xi_i^{-r_i} \right) \left(\prod_{j=1}^{n} \eta_j^{-c_j} \right) \left(\prod_{ij} \frac{1}{1 - \xi_i \eta_j} \right).$$

More precisely,

$$\begin{aligned} \int_{\Delta} \phi(X) \ dX \ \geq \ & \frac{\Gamma\left(\frac{m+n}{2}\right)}{2e^5 \pi^{\frac{m+n-2}{2}} \sqrt{mn}} \left(\frac{2}{(mn)^2 (N+1)(N+mn)} \right)^{m+n-1} \\ & \times \frac{(mn)^{mn}}{(mn)!} \frac{1}{(N+mn)^{N+mn}} \\ & \times \left(\prod_{i=1}^{m} \xi_i^{-r_i} \right) \left(\prod_{j=1}^{n} \eta_j^{-c_j} \right) \left(\prod_{ij} \frac{1}{1 - \xi_i \eta_j} \right) \end{aligned}$$

provided $m+n \geq 10$. Hence by (6.3.2) the number $\#(R,C)$ is estimated from below within a factor of $(N+mn)^{O(m+n)}$ by

$$\begin{aligned} e^N (N+mn)! \int_\Delta \phi(X)\ dX \ \approx & \frac{e^{mn+N}(N+mn)!}{(N+mn)^{N+mn}} \\ & \times \left(\prod_{i=1}^m \xi_i^{-r_i}\right) \left(\prod_{j=1}^n \eta_j^{-c_j}\right) \left(\prod_{ij} \frac{1}{1-\xi_i\eta_j}\right) \\ \approx & F(x^*, y^*) = \rho, \end{aligned}$$

where "$\approx$" stands for an approximation within a $N^{O(m+n)}$ factor.

More precisely, by (6.3.1)

$$\begin{aligned} \#(R,C) \ \geq \ & \frac{\Gamma\left(\frac{m+n}{2}\right)}{2e^5 \pi^{\frac{m+n-2}{2}} mn(N+mn)} \left(\frac{2}{(mn)^2(N+1)(N+mn)}\right)^{m+n-1} \\ & \times \left(\prod_{i=1}^m \frac{r_i^{r_i}}{r_i!}\right) \left(\prod_{j=1}^n \frac{c_j^{c_j}}{c_j!}\right) \frac{N!(N+mn)!(mn)^{mn}}{N^N (N+mn)^{N+mn}(mn)!} \rho(R,C) \end{aligned}$$

provided $m+n \geq 10$. □

The proof of Theorem 1.3 is a straightforward modification of the proof of Theorem 1.1.

(7.4) Proof of Theorem 1.3. The upper bound follows from the generating function expression

$$\prod_{ij} \frac{1}{1-w_{ij}x_i y_j} = \sum_{R,C} T(R,C;W) x^R y^C.$$

Let us prove the lower bound. Since $T(R,C;W)$ is a polynomial in W, without loss of generality we assume that W is a strictly positive matrix. Let

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

be the minimum point of

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

see Lemma 1.4. Then

$$\begin{aligned} & \sum_{j=1}^n \frac{w_{ij}\xi_i\eta_j}{1-w_{ij}\xi_i\eta_j} = r_i \quad \text{for} \quad i = 1, \ldots, m \\ & \sum_{i=1}^m \frac{w_{ij}\xi_i\eta_j}{1-w_{ij}\xi_i\eta_j} = c_j \quad \text{for} \quad j = 1, \ldots, n. \end{aligned} \tag{7.4.1}$$

In the space of matrices, let us consider the standard simplex Δ and the matrix $A = (\alpha_{ij})$

$$\alpha_{ij} = \frac{1}{(N+mn)(1 - w_{ij}\xi_i\eta_j)} \quad \text{for all} \quad i, j.$$

As in the proof of Theorem 1.1, we check from (7.4.1) that indeed $A \in \Delta$. Let

$$\lambda_i = \frac{1}{\xi_i\sqrt{N+mn}} \quad \text{for} \quad i = 1, \ldots, m \quad \text{and}$$
$$\mu_j = \frac{1}{\eta_j\sqrt{N+mn}} \quad \text{for} \quad j = 1, \ldots, n.$$

Let us consider the affine space $L \subset \mathbb{R}^d$ defined by the equations

$$\sum_{j=1}^{n} \frac{w_{ij}x_{ij}}{\lambda_i\mu_j} = r_i \quad \text{for} \quad i = 1, \ldots, m$$
$$\sum_{i=1}^{m} \frac{w_{ij}x_{ij}}{\lambda_i\mu_j} = c_j \quad \text{for} \quad j = 1, \ldots, n.$$

Then $A \in L$, the value of $\phi_{R,C;W}$ on $\Delta \cap L$ is constant and equal to

$$\tau = \frac{1}{(N+mn)^N} \left(\prod_{i=1}^{m} \xi_i^{-r_i}\right) \left(\prod_{j=1}^{n} \eta_j^{-c_j}\right),$$

see (6.5.2)-(6.5.3). Next, we use the lower bound in (6.5.1) and the proof proceeds as for Theorem 1.1. □

Acknowledgments

This work grew out of a joint project with Alex Samorodnitsky, and at an earlier stage also with Alexander Yong, on constructing computationally efficient algorithms for enumeration of contingency tables, see [BSY07]. I am grateful to Keith Ball and Matthieu Fradelizi for teaching me methods to estimate volumes of sections of convex bodies. I benefitted from conversations with Alex Samorodnitsky, Imre Bárány, and Roy Meshulam.

References

[BLV04] W. Baldoni-Silva, J.A. De Loera, and M. Vergne, *Counting integer flows in networks*, Found. Comput. Math. **4** (2004), 277–314.

[Bal88] K. Ball, *Logarithmically concave functions and sections of convex sets in R^n*, Studia Math. **88** (1988), 69–84.

[Bal97] K. Ball, *An elementary introduction to modern convex geometry*, Flavors of Geometry, Math. Sci. Res. Inst. Publ., 31, Cambridge Univ. Press, Cambridge, 1997, pp. 1–58.

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

[Ba08] A. Barvinok, *Enumerating contingency tables via random permanents*, Combin. Probab. Comput. **17** (2008), 1–19.

[BSY07] A. Barvinok, A, Samorodnitsky, and A. Yong, *Counting magic squares in quasi-polynomial time*, preprint, arXiv math.CO/0703227 (2007).

[BP03] M. Beck and D. Pixton, *The Ehrhart polynomial of the Birkhoff polytope*, Discrete Comput. Geom. **30** (2003), 623–637.

[BBK72] A. Békéssy, P. Békéssy, and J. Komlós, *Asymptotic enumeration of regular matrices*, Studia Sci. Math. Hungar. **7** (1972), 343–353.

[Ben74] E. Bender, *The asymptotic number of non-negative integer matrices with given row and column sums*, Discrete Math. **10** (1974), 217–223.

[Bol97] B. Bollobás, *Volume estimates and rapid mixing*, Flavors of Geometry, Math. Sci. Res. Inst. Publ., 31, Cambridge Univ. Press, Cambridge, 1997, pp. 151–182.

[Br73] L.M. Bregman, *Certain properties of nonnegative matrices and their permanents*, Dokl. Akad. Nauk SSSR **211** (1973), 27–30.

[CM07a] E.R. Canfield and B.D. McKay, *Asymptotic enumeration of contingency tables with constant margins*, preprint arXiv math.CO/0703600 (2007).

[CM07b] E.R. Canfield and B.D. McKay, *The asymptotic volume of the Birkhoff polytope*, preprint arXiv:0705.2422 (2007).

[C+05] Y. Chen, P. Diaconis, S.P. Holmes, and J.S. Liu, *Sequential Monte Carlo methods for statistical analysis of tables*, J. Amer. Statist. Assoc. **100** (2005), 109–120.

[CD03] M. Cryan and M. Dyer, *A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant*, Special issue on STOC2002 (Montreal, QC), J. Comput. System Sci. **67** (2003), 291–310.

[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*, Ann. Statist. **13** (1985), 845–913.

[DG04] P. Diaconis and A. Gamburd, *Random matrices, magic squares and matching polynomials*, Research Paper 2, 26 pp. (electronic), Electron. J. Combin. **11** (2004/06).

[DG95] P. Diaconis and A. Gangolli, *Rectangular arrays with fixed margins*, Discrete probability and algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., 72, Springer, New York, 1995, pp. 15–41.

[DLY07] J.A. De Loera, F. Liu, and R. Yoshida, *Formulas for the volumes of the polytope of doubly-stochastic matrices and its faces*, preprint arXiv math.CO/0701866 (2007).

[DKM97] M. Dyer, R. Kannan, and J. Mount, *Sampling contingency tables*, Random Structures Algorithms **10** (1997), 487–506.

[Eg81] G.P. Egorychev, *The solution of van der Waerden's problem for permanents*, Adv. in Math. **42** (1981), 299–305.

[Fa81] D.I. Falikman, *Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix (Russian)*, Mat. Zametki **29** (1981), 931–938.

[Frad97] M. Fradelizi, *Sections of convex bodies through their centroid*, Arch. Math. (Basel) **69** (1997), 515–522.

[Goo76] I.J. Good, *On the application of symmetric Dirichlet distributions and their mixtures to contingency tables*, Ann. Statist. **4** (1976), 1159–1189.

[GM07] C. Greenhill and B.D. McKay, *Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums*, preprint arXiv:0707.0340 (2007).

[Grü60] B. Grünbaum, *Partitions of mass-distributions and of convex bodies by hyperplanes*, Pacific J. Math. **10** (1960), 1257–1261.

[Khi57] A.I. Khinchin, *Mathematical Foundations of Information Theory*, Translated by R. A. Silverman and M. D. Friedman, 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.

[MO68] A.W. Marshall and I. Olkin, *Scaling of matrices to achieve specified row and column sums*, Numer. Math. **12** (1968), 83–90.

[Mor02] B.J. Morris, *Improved bounds for sampling contingency tables*, Random Structures Algorithms **21** (2002), 135–146.

[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.

[Pak00] I. Pak, *Four questions on Birkhoff polytope*, Ann. Comb. **4** (2000), 83–90.

[RS89] U.G. Rothblum and H. Schneider, *Scalings of matrices which have prespecified row sums and column sums via optimization*, Linear Algebra Appl. **114/115** (1989), 737–764.

[Sch92] M. Schmuckenschläger, *On the volume of the double stochastic matrices*, Acta Math. Univ. Comenian. (N.S.) **61** (1992), 189–192.

[Si64] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist. **35** (1964), 876–879.

[Vaa79] J.D. Vaaler, *A geometric inequality with applications to linear forms*, Pacific J. Math. **83** (1979), 543–553.

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

E-mail address: barvinok@umich.edu