An Approximation Algorithm for Counting Contingency Tables*

Alexander Barvinok,[1] **Zur Luria,**[2] **Alex Samorodnitsky,**[2] **Alexander Yong**[3]

[1]*Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109-1043; e-mail: barvinok@umich.edu*

[2]*Department of Computer Science, Hebrew University of Jerusalem, Givat Ram Campus, 91904, Israel; e-mail: zluria@cs.huji.ac.il; salex@cs.huji.ac.il*

[3]*Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801; e-mail: ayong@illinois.edu*

ABSTRACT: We present a randomized approximation algorithm for counting *contingency tables*, $m \times n$ non-negative integer matrices with given row sums $R = (r_1, \ldots, r_m)$ and column sums $C = (c_1, \ldots, c_n)$. We define *smooth margins* (R, C) in terms of the *typical table* and prove that for such margins the algorithm has quasi-polynomial $N^{O(\ln N)}$ complexity, where $N = r_1 + \cdots + r_m = c_1 + \cdots + c_n$. Various classes of margins are smooth, e.g., when $m = O(n)$, $n = O(m)$ and the ratios between the largest and the smallest row sums as well as between the largest and the smallest column sums are strictly smaller than the golden ratio $(1 + \sqrt{5})/2 \approx 1.618$. The algorithm builds on Monte Carlo integration and sampling algorithms for log-concave densities, the matrix scaling algorithm, the permanent approximation algorithm, and an integral representation for the number of contingency tables. © 2010 Wiley Periodicals, Inc. Random Struct. Alg., 37, 25–66, 2010

Keywords: contingency tables; randomized algorithms; matrix scaling; permanent approximation

1. INTRODUCTION

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

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

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

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

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

Let $\#(R, C)$ denote the number of these contingency tables.

There is interest in the study of $\#(R, C)$, due to connections to statistics, combinatorics and representation theory, see, e.g., [3, 13, 14, 16–19, 25, 36] and the references therein. However, since enumerating $\#(R, C)$ is a #*P*-complete problem even for $m = 2$ [19], one does not expect to find polynomial-time algorithms (nor formulas) computing $\#(R, C)$ exactly. As a result, attention has turned to the open problem of efficiently estimating $\#(R, C)$.

We present a randomized algorithm for approximating $\#(R, C)$ within a prescribed relative error. Based on earlier numerical studies [8, 41], we conjecture that its complexity is polynomial in N. We provide further evidence for this hypothesis: we introduce "smooth margins" (R, C) where the entries of the *typical table* are not too large, and among $\{r_1, \ldots, r_m, c_1, \ldots, c_n\}$ there are no "outliers." Our main result is that smoothness implies a quasi-polynomial $N^{O(\log N)}$ complexity bound on the algorithm. More precisely, we approximate $\#(R, C)$ within relative error $\epsilon > 0$ using $(1/\epsilon)^{O(1)} N^{O(\ln N)}$ time in the unit cost model, provided $\epsilon \gg 2^{-m} + 2^{-n}$.[1]

The class of smooth margins captures a number of interesting subclasses. In particular, this work applies to the case of *magic squares* (where $m = n$ and $r_i = c_j = t$ for all i, j), extending [8]. More generally, smoothness includes the case when the ratios m/n and n/m are bounded by a constant fixed in advance while the ratios between the largest and the smallest row sums as well as between the largest and the smallest column sums are smaller than the golden ratio $(1 + \sqrt{5})/2 \approx 1.618$. These and others examples are explicated in Section 3. See Section 1.3 for comparisons to the literature.

1.1. An Outline of the Algorithm

Our algorithm builds on the technique of rapidly mixing Markov chains and, in particular, on efficient integration and sampling from log-concave densities, as developed in [2, 23, 24, 33] (see also [40] for a survey), the permanent approximation algorithm [29], the strongly polynomial time algorithm for matrix scaling [32], and the integral representation of $\#(R, C)$ from [6].

Let $\Delta = \Delta_{m \times n} \subset \mathbb{R}^{mn}$ be the open $(mn - 1)$-dimensional simplex of all $m \times n$ positive matrices $X = (x_{ij})$ such that

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

---

[1]If an exponentially small relative error $\epsilon = O(2^{-m} + 2^{-n})$ is desired, one has an exact dynamic programming algorithm with $N^{O(m+n)} = (1/\epsilon)^{O(\ln N)}$ quasi-polynomial complexity.

Let dX be Lebesgue measure on Δ normalized to the probability measure. An integral representation for $\#(R, C)$ was found in [6]:

$$\#(R, C) = \int_\Delta f(X) dX, \tag{1}$$

where $f : \Delta \longrightarrow \mathbb{R}_+$ is a certain continuous function that factors as

$$f = p\phi, \tag{2}$$

where

$$p(X) \geq 1 \quad \text{for all} \quad X \in \Delta$$

is a function that "does not vary much", and $\phi : \Delta \longrightarrow \mathbb{R}_+$ is continuous and *log-concave*, that is,

$$\phi(\alpha X + \beta Y) \geq \phi^\alpha(X)\phi^\beta(Y) \quad \text{for all} \quad X, Y \in \Delta \quad \text{and}$$
$$\text{for all} \quad \alpha, \beta \geq 0 \quad \text{such that} \quad \alpha + \beta = 1.$$

Full details about f and its factorization are reviewed in Section 2.

For any $X \in \Delta$, the values of $p(X)$ and $\phi(X)$ are computable in time polynomial in N. Given $\epsilon > 0$, the value of $p(X)$ can be computed, within relative error ϵ in time polynomial in $1/\epsilon$ and N, by a randomized algorithm of [29] (function $p(X)$ is the permanent of a certain non-negative $N \times N$ matrix $B(X)$ parameterized by $X \in \Delta$). The value of $\phi(X)$ can be computed, within relative error ϵ in time polynomial in $\ln(1/\epsilon)$ and N, by a deterministic algorithm of [32].

The central idea of this article is to define smooth margins (R, C) so that matrices $X \in \Delta$ with large values of $p(X)$ do not contribute much to the integral (1). Our main results, precisely stated in Section 3, are that for smooth margins, there is a *threshold* $\tau = N^{\delta \ln N}$ for some constant $\delta > 0$ (depending on the class of margins considered) such that if we define the truncation $\bar{p} : \Delta \longrightarrow \mathbb{R}_+$ by

$$\bar{p}(X) = \begin{cases} p(X) & \text{if} \quad p(X) \leq \tau \\ \tau & \text{if} \quad p(X) > \tau \end{cases}$$

then

$$\#(R, C) = \int_\Delta p(X)\phi(X) \, dX \approx \int_\Delta \bar{p}(X)\phi(X) \, dX \tag{3}$$

where "$\approx$" means "approximates to within an $O(2^{-n} + 2^{-m})$ relative error" (in fact, rather than base 2, any constant $M > 1$, fixed in advance, can be used). We conjecture that one can choose the threshold $\tau = N^{O(1)}$, which would make the complexity of our algorithm polynomial in N.

The first step (and a simplified version) of our algorithm computes the integral

$$\int_\Delta \phi(X) \, dX \tag{4}$$

using any of the aforementioned randomized polynomial time algorithms for *integrating* log-concave densities; these results imply that this step has polynomial in N complexity. By

(3) it follows that for smooth (R, C) the integral (4) approximates $\#(R, C)$ within a factor of $N^{O(\ln N)}$. This simplified algorithm is suggested in [6]; an implementation that utilizes a version of the hit-and-run algorithm of [33], together with numerical results is described in [41] and [8].

Next, our algorithm estimates (3) within relative error ϵ using the aforementioned randomized polynomial time algorithm for approximating the permanent of a matrix, and any of those for *sampling* from log-concave densities. Specifically, let ν be the probability measure on Δ with the density proportional to ϕ. Thus,

$$\int_{\Delta} \bar{p}(X) \phi(X) \, \mathrm{d}X = \left(\int_{\Delta} \bar{p} \, \mathrm{d}\nu \right) \left(\int_{\Delta} \phi(X) \, \mathrm{d}X \right).$$

The second factor is computed by the above first step, while the first factor is approximated by the sample mean

$$\int_{\Delta} \bar{p} \, \mathrm{d}\nu \approx \frac{1}{k} \sum_{i=1}^{k} \bar{p}(X_i), \tag{5}$$

where $X_1, \ldots, X_k \in \Delta$ are independent points sampled at random from measure ν. Since $1 \leq \bar{p}(X) \leq \tau$, the Chebyshev inequality implies that to achieve relative error ϵ with probability 2/3 it suffices to sample $k = O(\epsilon^{-2}\tau^2) = \epsilon^{-2} N^{O(\ln N)}$ points in (5).

The results of [2, 23, 24, 33] imply that for any given $\epsilon > 0$ one can sample independent points $X_1, \ldots, X_k$ from a distribution $\tilde{\nu}$ on Δ such that

$$|\tilde{\nu}(S) - \nu(S)| \leq \epsilon \quad \text{for any Borel set} \quad S \subset \Delta$$

in time linear in k and polynomial in ϵ^{-1} and N. Replacing ν by $\tilde{\nu}$ in (5) introduces an additional relative error of $\epsilon \tau = \epsilon N^{\delta \ln N}$, handled by choosing a smaller $\epsilon = O(N^{-\delta \ln N})$.

1.2. An Optimization Problem, Typical Tables, and Smooth Margins

We will define smoothness of margins in terms of a certain convex optimization problem.

Let $\mathcal{P} = \mathcal{P}(R, C)$ be the *transportation polytope* of $m \times n$ non-negative matrices $X = (x_{ij})$ with row sums R and column sums C. On the space $\mathbb{R}_+^{mn}$ of $m \times n$ non-negative matrices define

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

The following optimization problem plays an important role in this paper:

$$\text{Maximize} \quad g(X) \quad \text{subject to} \quad X \in \mathcal{P}. \tag{6}$$

It is easy to check that g is strictly concave and hence attains its maximum on $\mathcal{P}$ at a unique matrix $X^* = (x_{ij}^*)$, $X^* \in \mathcal{P}$ that we call the *typical table*.

An intuitive explanation for the appearance of this optimization problem, and justification for the nomenclature "typical" derives from work of [7] (relevant parts are replicated for convenience, in Section 4, see specifically Theorem 4.1). In short, X^* determines the asymptotic behavior of $\#(R, C)$.

The main requirement that we demand of smooth margins (R, C) to satisfy (see Section 3 for unsuppressed technicalities) is that the entries of the typical table are not too large, that is, entries x_{ij}^* of the optimal solution $X^* = (x_{ij}^*)$ satisfy

$$\max_{ij} x_{ij}^* = O(s) \quad \text{where} \quad s = \frac{N}{mn}$$

is the average entry of the table.

As viewing the typical table is interesting in its own right, one would like to understand how the typical table changes as the margins vary. The optimization problem being convex, X^* can be computed efficiently by many existing algorithms, see, for example, [37]. However, in many instances of interest, the smoothness condition can be checked without actually needing to solve this problem. For example, if all the row sums r_i are equal, the symmetry of the functional g under permutations of rows implies that

$$x_{ij}^* = \frac{c_j}{m} \quad \text{for all} \quad i, j.$$

In general, the entries x_{ij}^* stay small if the row sums r_i and column sums c_j do not vary much. On the other hand, it is not hard to construct examples of margins (R, C) for n-vectors R and C such that $n \le r_i, c_j \le 3n$ and some of the entries x_{ij}^* are large, in fact linear in n. Another one of our results (Theorem 3.3) gives upper and lower bounds for x_{ij}^* in terms of (R, C).

1.3. Comparisons with the Literature

Using the Markov Chain Monte Carlo approach, Dyer, Kannan, and Mount [19] count contingency tables when R and C are sufficiently large, that is, if $r_i = \Omega(n^2 m)$ and $c_j = \Omega(m^2 n)$ for all i, j. Their randomized (sampling) algorithm approximates $\#(R, C)$ within any given relative error $\epsilon > 0$ in time polynomial in ϵ^{-1}, n, m, and $\sum_i \log r_i + \sum_j \log c_j$ (the bit size of the margins). Subsequently, Morris [36] obtained a similar result for the bounds $r_i = \Omega(n^{3/2} m \ln m)$ and $c_j = \Omega(m^{3/2} n \ln n)$. These results are based on fact that for large margins, the number of contingency tables is well-approximated by the volume of the transportation polytope $\mathcal{P}(R, C)$ (contingency tables being the integer points in this polytope). More generally, Kannan and Vempala [31] show that estimating the number integer points in a d-dimensional polytope with m facets reduces to computing the volume of the polytope (a problem, for which efficient randomized algorithms exist, see [40] for a survey) provided the polytope contains a ball of radius $d\sqrt{\log m}$.

When the margins r_i, c_j are very small, that is, bounded by a constant fixed in advance) relative to the sizes m and n of the matrix, Békéssy, Békéssy, and Komlós [9] obtain an efficient and precise asymptotic formula for $\#(R, C)$. Their formula exploits the fact in this case, the majority of contingency tables have only entries 0, 1, and 2. Alternatively, in this case one can exactly compute $\#(R, C)$ in time polynomial in $m + n$ via a dynamic programming algorithm. More recently, Greenhill and McKay [26] gave a computationally efficient asymptotic formula for a wider class of sparse margins (when $r_i c_j = o(N^{2/3})$).

Also using the dynamic programming approach, Cryan and Dyer [14] construct a randomized polynomial time approximation algorithm to compute $\#(R, C)$, provided the number of rows is fixed; see [15] for sharpening of the results.

It seems that the most resilient case of computing $\#(R, C)$ is where both m and n grow, and the margins are of moderate size, e.g., linear in the dimension. Recently, Canfield and

McKay [12] found a precise asymptotic formula for $\#(R, C)$ assuming that all row sums are equal and all column sums are equal. However, for general margins no such formula is known, even conjecturally. *Added in proofs:* an asymptotic formula for certain smooth classes of margins is found in A. Barvinok and J.A. Hartigan, An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums, preprint arXiv:0910.2477 (2009).

We remark that our notion of smooth margins includes all of the above regimes, except for that of large margins.

Summarizing, although our complexity bounds do not improve on the algorithms in the above cases, our algorithm is provably computationally efficient (quasi-polynomial in N) for several new classes of margins, which include cases of growing dimensions m and n and moderate size margins R and C.

2. THE INTEGRAL REPRESENTATION FOR THE NUMBER OF CONTINGENCY TABLES

We now give details of the integral representation (1). To do this, we express $\#(R, C)$ as the expectation of the permanent of a random $N \times N$ matrix. Recall that the *permanent* of an $N \times N$ matrix A is defined by

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

where S_N is the symmetric group of the permutations of the set $\{1, \ldots, N\}$.

It turns out that the permanent of a particularly structured matrix is closely related to the enumeration of contingency tables. The following result was proved in [6], see Theorem 1.1 there.

Theorem 2.1.   *For an $m \times n$ matrix $X = (x_{ij})$, let $A(X)$ be the $N \times N$ block matrix $A(X)$ whose (i, j)-th block is the $r_i \times c_j$ submatrix filled with x_{ij}, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Then*

$$\frac{\text{per } A(X)}{r_1! \cdots r_m! c_1! \cdots c_n!} = \sum_{D=(d_{ij})} \prod_{ij} \frac{x_{ij}^{d_{ij}}}{d_{ij}!}, \tag{7}$$

where the sum is over all non-negative integer matrices $D = (d_{ij})$ with row sums R and column sums C.

Let $\mathbb{R}_+^{mn}$ be the open orthant of positive $m \times n$ matrices X. Then

$$\#(R, C) = \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \int_{\mathbb{R}_+^{mn}} \text{per } A(X) \, \exp\left\{-\sum_{ij} x_{ij}\right\} \, dX, \tag{8}$$

where dX is the Lebesgue measure on $\mathbb{R}_+^{mn}$.

In the case that $r_i = a$ and $c_j = b$ for all i, j, the expansion (7) was first observed by Bang and then used by Friedland [22] in his proof of a weaker form of the van der Waerden conjecture; see Section 7.1 and references there. As an anonymous referee pointed out to

us, a similar in spirit to (7) formula was observed by Bender (formula (3) in [10]) and other similar combinatorial identities have been used since then. The gist of Theorem 2.1 is the integral representation (8) which allows us to represent a discrete quantity (the number of integer matrices) as an integral of a positive continuous function which turns out to possess certain desirable analytical and computational properties.

Since the function $X \longmapsto \operatorname{per} A(X)$ is a homogeneous polynomial of degree N, one can express $\#(R, C)$ as an integral over the simplex. The following corollary was also obtained in [6].

Corollary 2.2. *Let* $\Delta = \Delta_{m \times n} \subset \mathbb{R}^{mn}$ *be the open simplex of positive* $m \times n$ *matrices* $X = (x_{ij})$ *such that* $\sum_{ij} x_{ij} = 1$. *Then*

$$\#(R, C) = \frac{(N + mn - 1)!}{(mn - 1)!} \frac{1}{r_1! \ldots r_m! c_1! \cdots c_n!} \int_{\Delta_{m \times n}} \operatorname{per} A(X) \, dX,$$

where dX *is the Lebesgue measure on* $\Delta_{m \times n}$ *normalized to the probability measure.*

Hence in the integral representation (1), we define the function f by

$$f(X) = \frac{(N + mn - 1)!}{(mn - 1)!} \frac{1}{r_1! \ldots r_m! c_1! \cdots c_n!} \operatorname{per} A(X)$$

$$= \frac{(N + mn - 1)!}{(mn - 1)!} \sum_{D=(d_{ij})} \prod_{ij} \frac{x_{ij}^{d_{ij}}}{d_{ij}!},$$

where $A(X)$ is the block matrix of Theorem 2.1 and the sum is over all contingency tables D with margins (R, C).

2.1. Matrix Scaling and the Factorization of *f*

To obtain the factorization (2), where $\phi : \Delta \longrightarrow \mathbb{R}_+$ is a log-concave function and $p : \Delta \longrightarrow \mathbb{R}_+$ is a function which "does not vary much", we employ the idea of *matrix scaling*, see [30, 35, 38], Chapter 6 of [4], and [32]: Let $X = (x_{ij})$ be a positive $m \times n$ matrix. Then there exists a unique $m \times n$ matrix Y with the row sums $R = (r_1, \ldots, r_m)$, column sums $C = (c_1, \ldots, c_n)$, and such that

$$x_{ij} = y_{ij} \lambda_i \mu_j \quad \text{for all} \quad i, j$$

and some positive $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n$. The numbers λ_i and μ_j are unique up to a rescaling $\lambda_i \longmapsto \lambda_i \tau$, $\mu_j \longmapsto \mu_j \tau^{-1}$. Note that if we divide the entries in the (i, j)-th block of the matrix $A(X)$ of Theorem 2.1 by $r_i c_j \lambda_i \mu_j$, we obtain a positive *doubly stochastic* matrix $B(X)$, that is, a positive matrix with all row and column sums equal to 1. Thus we have

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

It is proved in [6] that

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

The lower bound is the van der Waerden bound for permanents of doubly stochastic matrices, see [21], [20] and also Chapter 12 of [34] and recent [28], while the upper bound is a corollary of the Minc conjecture proved by Bregman, see [11], Chapter 11 of [34], and also [39].

Now we define

$$p(X) = \frac{N^N}{N!} \operatorname{per} B(X) \tag{9}$$

and

$$\phi(X) = \frac{(N + mn - 1)!N!}{(mn - 1)!N^N} \left(\prod_{i=1}^m \frac{r_i^{r_i}}{r_i!} \right) \left(\prod_{j=1}^n \frac{c_j^{c_j}}{c_j!} \right) \left(\prod_{i=1}^m \lambda_i^{r_i} \right) \left(\prod_{j=1}^n \mu_j^{c_j} \right).$$

We summarize results of [6] regarding p and ϕ.

Theorem 2.3. *The following hold:*

1. *ϕ is log-concave, that is,*

$$\phi(\alpha X + \beta Y) \geq \phi^\alpha(X)\phi^\beta(Y)$$

   for all $X, Y \in \Delta$ and $\alpha, \beta \geq 0$ such that $\alpha + \beta = 1$;
2. *Let $X, Y \in \Delta$ be positive $m \times n$ matrices, $X = (x_{ij})$ and $Y = (y_{ij})$, such that $x_{ij}, y_{ij} \geq \delta$ for all i, j and some $\delta > 0$. Then*

$$|\ln \phi(X) - \ln \phi(Y)| \leq \frac{N}{\delta} \max_{ij} |x_{ij} - y_{ij}|;$$

3. *For $\delta < 1/mn$ let us define the δ-interior Δ_δ of the simplex Δ as the set of matrices $X \in \Delta$, $X = (x_{ij})$, such that $x_{ij} \geq \delta$ for all i, j. Then for $f = p\phi$ we have*

$$(1 - mn\delta)^{N+mn-1} \int_\Delta f \, dX \leq \int_{\Delta_\delta} f \, dX \leq \int_\Delta f \, dX;$$

4. *We have*

$$1 \leq p(X) \leq \frac{N^N}{N!} \min \left\{ \prod_{i=1}^m \frac{r_i!}{r_i^{r_i}}, \prod_{j=1}^n \frac{c_j!}{c_j^{c_j}} \right\}.$$

The log-concavity of function ϕ was first observed in [27]. In terms of [27], up to a normalization factor, $\phi(X)$ is the *capacity* of the matrix $A(X)$ of Theorem 2.1, see also [5] for a more general family of inequalities satisfied by ϕ. As is discussed in [6], the matrix scaling algorithm of [32] leads to a polynomial time algorithm for computing $\phi(X)$. Namely, for any given $\epsilon > 0$ the value of $\phi(X)$ can be computed within relative error of ϵ in time polynomial in N and $\ln(1/\epsilon)$ in the unit cost model; our own experience is that this algorithm for computing $\phi(X)$ is practical, and works well for $m, n \leq 100$.

Theorems 2.3 and 2.1 allow us to apply algorithms of [2, 23, 24, 33] on efficient integration and sampling of log-concave functions. First, for any given $\epsilon > 1$, one can compute the integral

$$\int_\Delta \phi \, dX$$

within relative error ϵ in time polynomial in ϵ^{-1} and N by a randomized algorithm. Second, one can sample points $X_1, \ldots, X_k \in \Delta$ independently from a measure $\tilde{\nu}$ such that

$$|\nu(S) - \tilde{\nu}(S)| \leq \epsilon \quad \text{for any Borel set} \quad S \subset \Delta,$$

where ν is the measure with the density proportional to ϕ, in time polynomial in k, ϵ^{-1} and N.

The integration of $p(X)$ raises a greater challenge. For any given $\epsilon > 0$ one can compute $p(X)$ itself within relative error ϵ in time polynomial in ϵ^{-1} and N, using the permanent approximation algorithm of [29]. However, the upper bound of Part (4) of Theorem 2.3 is, in the worst case, of order $N^{\gamma(m+n)}$ for some absolute constant $\gamma > 0$. Therefore, a priori, to integrate p over Δ using a sample mean, one needs too many such computations to guarantee the desired accuracy of ϵ. Our main observation to overcome this problem is that in many interesting cases the matrices $X \in \Delta$ with large values of $p(X)$ do not contribute much to the integral (1), so we have $p(X) = N^{O(\ln N)}$ with high probability with respect to the density on Δ proportional to f.

2.2. Bounding p with High Probability

To bound probabilities of various events in the simplex Δ with respect to the density proportional to f, it is convenient to extend p from simplex Δ to the non-negative orthant $\mathbb{R}_+^{mn}$ and estimate probabilities of events in the orthant. Our main idea is to show that with respect to the probability density $\psi = \psi_{R,C}$ on $\mathbb{R}_+^{mn}$ that is proportional to the integrand per $A(X) \exp\left\{ -\sum_{ij} x_{ij} \right\}$ in (8), the overwhelming majority of matrices $X \in \mathbb{R}_+^{mn}$ are such that the permanent of the doubly stochastic scaling $B(X)$ of $A(X)$ is close to its lower bound $N!/N^N$.

More formally, we consider the projection which scales every positive matrix to the total sum 1 of all entries:

$$\text{pr} : \mathbb{R}_+^{mn} \longrightarrow \Delta_{m \times n}, \quad pr(X) = \tilde{X}, \quad \text{where}$$

$$\tilde{X} = \alpha X \quad \text{for} \quad \alpha = \left(\sum_{ij} x_{ij} \right)^{-1}.$$

Since projection pr just multiplies a matrix by a number, the scalings of X and $\tilde{X}$ to the matrix with the row sums R and column sums C coincide. Also, the doubly stochastic scalings $B(X)$ and $B(\tilde{X})$, of matrices $A(X)$ and $A(\tilde{X})$, respectively, coincide. We define $p(X)$ for an arbitrary positive $m \times n$ matrix X by $p(X) := p(\tilde{X})$, or, equivalently, by (9).

Theorem 2.1 gives us the following formula for the probability density $\psi = \psi_{R,C}$ on $\mathbb{R}_+^{mn}$ proportional to the integrand in (8):

$$\psi(X) = \frac{1}{\#(R,C)} \sum_{D=(d_{ij})} \prod_{ij} \frac{x_{ij}^{d_{ij}}}{d_{ij}!} e^{-x_{ij}}, \quad \text{where}$$

$$X = (x_{ij}) \quad \text{and} \quad x_{ij} > 0 \quad \text{for all} \quad i, j,$$

and the sum is over all $m \times n$ non-negative integer matrices D with the row sums R and column sums C. We define $\psi(X) = 0$ if X is not a positive matrix.

Our goal is to show that for smooth margins (R, C), the value of $p(X)$ is "reasonably small" for most X, that is,

$$\mathbf{P}\left\{X \in \mathbb{R}_+^{mn} : p(X) > N^{\delta \ln N}\right\} < \kappa(2^{-m} + 2^{-n}) \tag{10}$$

for some constants $\delta > 0$ and $\kappa > 0$, where the probability is measured with respect to the density ψ. Then the projection pr makes sure that the value of $p(X)$ is "reasonably small" for most $X \in \Delta$ with respect to the density on the simplex Δ proportional to f.

Our construction of function f in (1) implies that the push-forward of ψ under the projection $\mathrm{pr} : \mathbb{R}_+^{mn} \longrightarrow \Delta$ is the density

$$\frac{1}{\#(R, C)}f(X) \quad \text{for} \quad X \in \Delta$$

on the simplex. Hence inequality (10) implies that for $\tau = N^{\delta \ln N}$ we have

$$\frac{1}{\#(R, C)} \int\limits_{\substack{X \in \Delta \\ p(X) > \tau}} f(X)\, \mathrm{d}X < \kappa(2^{-m} + 2^{-n}).$$

Therefore, as discussed in Section 1.1, replacing p by its truncation $\bar{p}$ introduces an $O(2^{-n} + 2^{-m})$ relative error in (3) and hence our algorithm achieves quasi-polynomial complexity.

The key idea behind inequality (10) is that the permanent of an appropriately defined "random" doubly stochastic matrix is very close with high probability to the van der Waerden lower bound $N!/N^N$; see Lemma 5.1.

3. MAIN RESULTS

Now we are ready to precisely define the classes of smooth margins for which our algorithm achieves $N^{O(\ln N)}$ complexity.

3.1. Smoothness Definitions

Fix margins $R = (r_1, \ldots, r_m)$, $C = (c_1, \ldots, c_n)$, where

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

Let

$$s = \frac{N}{mn}$$

be the average value of the entries of the table. We define

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

Hence r_+ and c_+ are the largest row and column sums, respectively, and r_- and c_- are the smallest row and column sums, respectively.

For $s_0 > 0$, call the margins (R, C) s_0-*moderate* if $s \le s_0$. In other words, margins are moderate if the average entry of the table is bounded from above.

For $\alpha \ge 1$, the margins (R, C) are *upper α-smooth* if

$$r_+ \le \alpha s n = \alpha \frac{N}{m} \quad \text{and} \quad c_+ \le \alpha s m = \alpha \frac{N}{n}.$$

Thus, margins are upper smooth if the row and column sums are at most proportional to the average row and column sums, respectively.

For $0 < \beta \le 1$, the margins (R, C) are *lower β-smooth* if

$$r_- \ge \beta s n = \beta \frac{N}{m} \quad \text{and} \quad c_- \ge \beta s m = \beta \frac{N}{n}.$$

Therefore, margins are lower smooth if the row and column sums are at least proportional to the average row and column sums, respectively.

The key smoothness condition is as follows: for $\alpha \ge 1$ we define margins (R, C) to be *strongly upper α-smooth* if for the typical table $X^* = (x_{ij}^*)$ we have

$$x_{ij}^* \le \alpha s = \alpha \frac{N}{mn} \quad \text{for all} \quad i, j.$$

Note that this latter condition implies that the margins are upper α-smooth. (Also, we do not need a notion of strongly lower β smooth.)

Our main results are randomized approximation algorithms of quasi-polynomial $N^{O(\ln N)}$ complexity when the margins (R, C) are smooth for either:

- s_0-moderate strongly upper α-smooth, for some fixed s_0 and α; or
- lower β and strongly upper α-smooth, for some fixed α and β.

By the discussion of Section 2.2, the quasi-polynomial complexity claim about our algorithm follows from bounding on $p(X)$ with high probability. Specifically, we have the following two results. Their proofs are argued similarly, but the second is more technically involved.

Theorem 3.1. *Fix $s_0 > 0$ and $\alpha \ge 1$. Suppose that $m \le 2^n$, $n \le 2^m$ and let (R, C) be s_0-moderate strongly upper α-smooth margins. Let $X = (x_{ij})$ be a random $m \times n$ matrix with density ψ of Section 2.2 , and let $p : \mathbb{R}_+^{mn} \longrightarrow \mathbb{R}_+$ be the function defined in Section 2.1. Then for some constant $\delta = \delta(\alpha, s_0) > 0$ and some absolute constant $\kappa > 0$, we have*

$$\mathbf{P}\{X : \quad p(X) > N^{\delta \ln N}\} \le \kappa (2^{-m} + 2^{-n}).$$

Therefore, the algorithm of Section 1.1 achieves $N^{O(\ln N)}$ complexity on these classes of margins. More precisely, for any $\epsilon > 2\kappa(2^{-m} + 2^{-n})$ the algorithm computes the number $\#(R, C)$ of tables within relative error ϵ in $(1/\epsilon)^{O(1)} N^{O(\ln N)}$ time.

Theorem 3.2. *Fix $\alpha \ge 1$, $0 < \beta \le 1$, and $\rho \ge 1$. Suppose that $m \le \rho n$, $n \le \rho m$ and let (R, C) be lower β and strongly upper α-smooth margins. Let $X = (x_{ij})$ be a random $m \times n$*

matrix with density ψ of Section 2.2 and let $p : \mathbb{R}_+^{mn} \longrightarrow \mathbb{R}_+$ be the function defined in Section 2.1. Then for some constant $\delta = \delta(\rho, \alpha, \beta) > 0$ and some absolute constant $\kappa > 0$, we have

$$\mathbf{P}\{X : \quad p(X) > N^{\delta \ln N}\} \leq \kappa(2^{-m} + 2^{-n}).$$

Therefore, the algorithm of Section 1.1 achieves $N^{O(\ln N)}$ complexity on these classes of margins. More precisely, for any $\epsilon > 2\kappa(2^{-m} + 2^{-n})$ the algorithm computes the number #(R, C) of tables within relative error ϵ in $(1/\epsilon)^{O(1)}N^{O(\ln N)}$ time.

We remark that in Theorem 3.1 and Theorem 3.2 above, we can replace base 2 by any base $M > 1$, fixed in advance.

3.2. Example: Symmetric Margins

While conditions for r_+, c_+, r_-, and c_- are straightforward to verify, to check the upper bounds for x_{ij}^* one may have to solve the optimization problem (6) first. There are, however, some interesting cases where an upper bound on x_{ij}^* can be inferred from symmetry considerations.

Note that if two row sums r_{i_1} and r_{i_2} are equal then the transportation polytope $\mathcal{P}(R, C)$ is invariant under the transformation which swaps the i_1-st and i_2-nd rows of a matrix $X \in \mathcal{P}(R, C)$. Since the function g in the optimization problem (6) also remains invariant if the rows are swapped and is strictly concave, we must have $x_{i_1 j}^* = x_{i_2 j}^*$ for all j. Similarly, if $c_{j_1} = c_{j_2}$ we must have $x_{ij_1}^* = x_{ij_2}^*$ for all i. In particular, if all row sums are equal, we must have $x_{ij}^* = c_j/m$. Similarly, if all column sums are equal, we must have $x_{ij}^* = r_i/n$.

More generally, one can show (see the proof of Theorem 3.3 in Section 6) that the largest entry x_{ij}^* of X^* necessarily lies at the intersection of the row with the largest row sum r_+ and the column with the largest column sum c_+. Therefore, if k of the row sums r_i are equal to r_+ we must have $x_{ij}^* \leq c_+/k$. Similarly, if k of the column sums are equal to c_+, we must have $x_{ij}^* \leq r_+/k$.

Here are some examples of classes margins where our algorithm probably achieves an $N^{O(\ln N)}$ complexity.

- The class of margins for which at least a constant fraction of the row sums r_i are equal to r_+:

$$\#\{i : \ r_i = r_+\} = \Omega(m)$$

  while m, n, the row, and the column sums differ by a factor, fixed in advance: $m/n = O(1)$, $n/m = O(1)$, $r_+/r_- = O(1)$, $c_+/c_- = O(1)$. Indeed, in this case we have

$$\max_{ij} x_{ij}^* = O(c_+/m) = O(N/mn)$$

  and quasi-polynomiality follows by Theorem 3.2.
- The class of margins for which at least a constant fraction of the row sums r_i are equal to r_+, while the column sums exceed the number of rows by at most a factor, fixed in advance, $c_+ = O(m)$, and m and n are not too disparate: $m \leq 2^n$ and $n \leq 2^m$. Indeed, in this case

$$\max_{ij} x_{ij}^* = O(c_+/m) = O(1)$$

and quasi-polynomiality follows by Theorem 3.1.
- The classes of margins defined as above, but with rows and columns swapped.

For a different source of examples, we prove that if both ratios r_+/r_- and c_+/c_- are not too large, the margins are strongly upper smooth. To do this, we use the following general result about the typical table X^*, to be proved in Section 6:

Theorem 3.3.　　*Let $X^* = (x_{ij}^*)$ be the typical table.*

　1. *We have*

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

　2. *If $r_- c_+ + r_- c_- + m r_- > r_+ c_+$ then*

$$x_{ij}^* \leq \frac{c_+(r_- c_- + m r_+)}{m(r_- c_+ + r_- c_- + m r_- - r_+ c_+)} \quad for \ all \quad i, j.$$

　　Similarly, if $c_- r_+ + c_- r_- + n c_- > r_+ c_+$ then

$$x_{ij}^* \leq \frac{r_+(c_- r_- + n c_+)}{n(c_- r_+ + c_- r_- + n c_- - c_+ r_+)} \quad for \ all \quad i, j.$$

3.3. Example: Golden Ratio Margins

Fix

$$1 \leq \beta < \frac{1 + \sqrt{5}}{2} \approx 1.618$$

and a number $\rho \geq 1$. Consider the class of margins (R, C) such that $m \leq \rho n$, $n \leq \rho m$, and

$$r_+/r_-, \ c_+/c_- \leq \beta.$$

We claim that our algorithm has an $N^{O(\ln N)}$ complexity on this class of margins.
　　To see this, let

$$\beta_1 = r_+/r_- \quad and \quad \beta_2 = c_+/c_-.$$

If $\beta_1 \leq \beta_2$ then

$$r_- c_+ + r_- c_- - r_+ c_+ = (1 + \beta_2 - \beta_1 \beta_2) r_- c_- \geq (1 + \beta_2 - \beta_2^2) r_- c_- \geq \epsilon r_- c_-$$

for some $\epsilon = \epsilon(\beta) > 0$ and hence by Part (2) of Theorem 3.3 we have

$$x_{ij}^* \leq \frac{c_+}{m} \left(\frac{1}{\epsilon} + \beta \right).$$

Similarly, if $\beta_2 \leq \beta_1$ then

$$c_- r_+ + c_- r_- - c_+ r_+ = (1 + \beta_1 - \beta_1 \beta_2) r_- c_- \geq (1 + \beta_1 - \beta_1^2) r_- c_- \geq \epsilon r_- c_-$$

for some $\epsilon = \epsilon(\beta) > 0$ and hence

$$x_{ij}^* \leq \frac{r_+}{n} \left(\frac{1}{\epsilon} + \beta \right).$$

In either case, (R, C) are strongly upper α-smooth for some $\alpha = \alpha(\beta)$ and Theorem 3.2 implies that our algorithm has a quasi-polynomial complexity on such margins. More generally, the algorithm is quasi-polynomial on the class of margins for which $\beta_1 = r_+/r_-$ and $\beta_2 = c_+/c_-$ are bounded above by a constant fixed in advance and $\beta_1\beta_2 \leq \max\{\beta_1, \beta_2\} + 1 - \epsilon$ where $\epsilon > 0$ is fixed in advance.

3.4. Example: Linear Margins

Fix $\beta \geq 1$ and $\epsilon > 0$ such that $\epsilon\beta < 1$ and consider the class of margins (R, C) for which

$$r_+/r_- \leq \beta \quad \text{and} \quad c_+ \leq \epsilon m.$$

Part (2) of Theorem 3.3 implies that the margins (R, C) are strongly upper α-smooth for some $\alpha = \alpha(\beta, \epsilon)$ and therefore quasi-polynomiality of the algorithm is guaranteed by Theorem 3.1.

The remainder of this paper is devoted to the proofs of Theorems 3.1, 3.2, and 3.3. While the proof of Theorem 3.3 is relatively straightforward, our proofs of Theorem 3.1 and especially Theorem 3.2 require some preparation. A general plan of the proofs of Theorems 3.1 and 3.2 is given in Section 5.

4. ASYMPTOTIC ESTIMATES

The following result is proved in [7] (see Theorem 1.1 and Lemma 1.4 there) and provides an asymptotic estimate for the number $\#(R, C)$ of contingency tables. It explains the role played by the optimization problem (6). It also introduces ingredients needed in the statement and proof of Theorem 5.3 given below.

Theorem 4.1. *Let $\mathcal{P}(R, C)$ be the transportation polytope of non-negative matrices with row sums R and column sums C and let $X^* = (x_{ij}^*)$ be the typical table, that is, the matrix $X^* \in \mathcal{P}(R, C)$ maximizing*

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

on $\mathcal{P}(R, C)$. Let

$$\rho(R, C) = \exp\{g(X^*)\} = \max_{\substack{X = (x_{ij}) \\ X \in \mathcal{P}(R,C)}} \prod_{ij} \frac{(x_{ij} + 1)^{x_{ij}+1}}{x_{ij}^{x_{ij}}}.$$

Then

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

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

Another representation of $\rho(R,C)$ is

$$\rho(R,C) = \min_{\substack{0 < x_1,\ldots,x_m < 1 \\ 0 < y_1,\ldots,y_n < 1}} \left(\prod_{i=1}^{m} x_i^{-r_i} \right) \left(\prod_{j=1}^{n} y_j^{-c_j} \right) \left(\prod_{ij} \frac{1}{1 - x_i y_j} \right).$$

A point $x_1, \ldots, x_m; y_1, \ldots, y_n$ minimizing the above product exists and is unique up to scaling $x_i \longmapsto x_i \tau,\ y_j \longmapsto y_j \tau^{-1}$. It is related to X^* by

$$x_{ij}^* = \frac{x_i y_j}{1 - x_i y_j} \quad \text{for all} \quad i,j.$$

We need the notion of the weighted enumeration of tables, as introduced in [6] and [5].

4.1. Weighted Enumeration of Tables

Fix margins R and C and a non-negative $m \times n$ matrix W. Define

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

where the sum is taken over all $m \times n$ non-negative integer matrices D with the row sums R and column sums C and we agree that $w_{ij}^0 = 1$. Therefore,

$$\#(R,C) = T(R,C;\mathbf{1}),$$

where $\mathbf{1}$ is the matrix of all 1's.

The estimates of Theorem 4.1 extend to weighted enumeration. We state only the part we are going to use. The following result is proved in [7] (see Theorem 1.3 there).

Theorem 4.2. *Let*

$$\rho(R,C;W) = \inf \left(\prod_{i=1}^{m} x_i^{-r_i} \right) \left(\prod_{j=1}^{n} y_j^{-c_j} \right) \left(\prod_{ij} \frac{1}{1 - w_{ij} x_i y_j} \right),$$

where the infimum is taken over all $x_1, \ldots, x_m > 0$ and $y_1, \ldots, y_n > 0$ such that $w_{ij} x_i y_j < 1$ for all i and j. Then

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

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

In fact, we will only use the upper bound of Theorem 4.2, which is actually straightforward to prove since $\prod_{ij} (1 - w_{ij} x_i y_j)^{-1}$ is the generating function for the family $T(R,C;W)$.

5. THE PLAN OF THE PROOFS OF THEOREMS 3.1 AND 3.2

To prove Theorems 3.1 and 3.2 we need to understand the behavior of the function

$$p(X) = \frac{N^N}{N!}\text{per } B(X),$$

that is, to estimate values of permanents of doubly stochastic matrices. The following straightforward corollary of results of [11, 20, 21, 39] shows that the permanent of an $N \times N$ doubly stochastic matrix lies close to $N!/N^N$ provided the entries of the matrix are not too large. We recall the definition of the Gamma function

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} \, dx \quad \text{for} \quad t > 0.$$

Lemma 5.1.   *Let $B = (b_{ij})$ be an $N \times N$ doubly stochastic matrix and let*

$$z_i = \max_{j=1,\dots,N} b_{ij} \quad for \quad i = 1,\dots,N.$$

Suppose that

$$\sum_{i=1}^N z_i \le \tau \quad for\ some \quad \tau \ge 1.$$

Then

$$\frac{N!}{N^N} \le \text{per } B \le \left(\frac{\tau}{N}\right)^N (\Gamma(1 + \tau^{-1}N))^\tau$$

$$\le \frac{N!}{N^N}(2\pi N)^{\tau/2} e^{\tau^2/12N}.$$

We delay the proof of Lemma 5.1 until Section 7.

We will apply Lemma 5.1 when $\tau = O(\ln N)$, in which case the ratio between the upper and lower bounds becomes $N^{O(\ln N)}$. In addition, we apply the lemma to the matrix $B(X)$, the doubly stochastic scaling of the random matrix $A(X)$ constructed in Theorem 2.1, see also Section 2.1. However, to use this lemma, we need to bound the entries of $B(X)$. To do that, we will need to be able to bound the entries of the matrix Y obtained from scaling X to have row sums R and column sums C. To this end, we prove the following result in Section 8, which might be of independent interest.

Theorem 5.2.   *Let $R = (r_1,\dots,r_m)$ and $C = (c_1,\dots,c_n)$ be positive vectors such that*

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

Let $X = (x_{ij})$ be an $m \times n$ positive matrix and let $Y = (y_{ij})$ be the scaling of X to have row sums R and column sums C, where

$$y_{ij} = \lambda_i \mu_j x_{ij} \quad for\ all \quad i,j$$

and some positive $\lambda_1,\dots,\lambda_m; \mu_1,\dots,\mu_n$.

Then, for every $1 \leq p \leq m$ and $1 \leq q \leq n$ we have

$$\ln y_{pq} \leq \ln \frac{r_p c_q}{N} + \ln x_{pq} + \ln \left(\frac{1}{N^2} \sum_{ij} r_i c_j x_{ij} \right) - \frac{1}{N} \sum_{j=1}^{n} c_j \ln x_{pj} - \frac{1}{N} \sum_{i=1}^{m} r_i \ln x_{iq}.$$

Now suppose that (R, C) are upper α-smooth margins, that is, $r_i/N \leq \alpha/m$ and $c_j/N \leq \alpha/n$ for some $\alpha \geq 1$, fixed in advance. To give an idea of the proofs of Theorems 3.1 and 3.2, suppose of the argument and the role of the hypotheses, suppose further that x_{ij} are sampled independently at random from the uniform distribution on $[0, 1]$. Then Theorem 5.2 and the law of large numbers clearly imply that as m and n grow, with overwhelming probability we have

$$y_{ij} \leq \kappa \frac{r_i c_j}{N} x_{ij} \quad \text{for all} \quad i, j$$

and some absolute constant $\kappa > 1$. If we construct the doubly stochastic matrix $B(X)$ as in Section 2.1, then with overwhelming probability for the entries b_{ij} we will have

$$b_{ij} \leq \frac{\kappa}{N} \quad \text{for all} \quad i, j.$$

However, in the situation of our proof, the matrix $X = (x_{ij})$ is actually sampled from the distribution with density ψ of Section 2.2. Thus to perform a similar analysis, we need to show that the entries of a random matrix X are uniformly small. For that, we have to assume that the margins (R, C) are *strongly* upper α-smooth (in fact, one can show that merely the condition of upper smoothness is not enough). Specifically, in Section 9, we prove the following result:

Theorem 5.3. *Let*

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

be a set of indices, and let $X = (x_{ij})$ be a random $m \times n$ matrix with density $\psi = \psi_{R,C}$ of Section 2.2. Suppose that the typical table $X^* = (x_{ij}^*)$ satisfies

$$x_{ij}^* \leq \lambda \quad \text{for all} \quad i, j$$

and some $\lambda > 0$.
Then for all $t > 0$ we have

$$\mathbf{P} \left\{ \sum_{(i,j) \in S} x_{ij} \geq t \right\} \leq \exp \left\{ -\frac{t}{2\lambda + 2} \right\} 4^{\#S} N^{\gamma(m+n)},$$

where $\gamma > 0$ is the absolute constant of Theorem 4.1.

We will use Theorem 5.3 to establish that a random matrix X is very unlikely to have many rows with large entries (Corollary 9.2). For that, we apply Theorem 5.3 to sets S having exactly one entry in each row of X.

In Section 10 we complete the proof of Theorem 3.1. Theorem 3.2 requires some more work and its proof is given in Section 12, after some technical estimates in Section 11.

6. PROOF OF THEOREM 3.3

First, we observe that the typical table $X^* = (x_{ij}^*)$ is strictly positive, that is, it lies in the interior of the transportation polytope $\mathcal{P}(R, C)$. Indeed, suppose that $x_{11}^* = 0$, for example. Choose indices p and q such that $x_{1q}^* > 0$ and $x_{p1}^* > 0$. Then necessarily $x_{pq}^* < r_p, c_q$ and we can consider a perturbation $X(\epsilon) \in \mathcal{P}(R, C)$ of X^* defined for sufficiently small $\epsilon > 0$ by

$$
x_{ij} = \begin{cases} x_{ij}^* + \epsilon & \text{if} \quad i = 1 & \text{and} \quad j = 1 \\ x_{ij}^* - \epsilon & \text{if} \quad i = p, j = 1 & \text{or} \quad i = 1, j = q \\ x_{ij}^* + \epsilon & \text{if} \quad i = p & \text{and} \quad j = q \\ x_{ij}^* & \text{if} \quad i \neq p & \text{and} \quad j \neq q. \end{cases}
$$

Since the value of

$$
\frac{\partial}{\partial x_{ij}} g(X) = \ln \left(\frac{x_{ij} + 1}{x_{ij}} \right)
$$

is equal to $+\infty$ at $x_{ij} = 0$ (we consider the right derivative in this case) and finite if $x_{ij} > 0$, we conclude that for a sufficiently small $\epsilon > 0$, the matrix $X(\epsilon)$ attains a larger value of $g(X)$, which is a contradiction. We conclude that all the entries of the typical table X^* are strictly positive.

Since X^* lies in the interior of the transportation polytope $\mathcal{P}(R, C)$, the Lagrange multiplier condition implies that

$$
\ln \left(\frac{x_{ij}^* + 1}{x_{ij}^*} \right) = \lambda_i + \mu_j \quad \text{for all} \quad i, j \tag{11}
$$

and some $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$. It follows that if $x_{i_1 j}^* \geq x_{i_2 j}^*$ for some row indices i_1, i_2 and *some* column index j then $\lambda_{i_1} \leq \lambda_{i_2}$ and hence $x_{i_1 j}^* \geq x_{i_2 j}^*$ for the same row indices i_1 and i_2 and *all* column indices j.

We prove Part (1) first. Let us choose a row i_0 with the largest row sum r_+. Without loss of generality, we assume that $i_0 = 1$. Hence

$$
x_{1j}^* \geq x_{ij}^* \quad \text{for} \quad j = 1, \ldots, n.
$$

Therefore,

$$
x_{1j}^* \geq \frac{c_j}{m} \geq \frac{c_-}{m} \quad \text{for} \quad j = 1, \ldots, n.
$$

Let us compare the entries in the first row and in the i-th row. From (11) we have

$$
\ln \left(\frac{x_{1j}^* + 1}{x_{1j}^*} \right) - \ln \left(\frac{x_{ij}^* + 1}{x_{ij}^*} \right) = \lambda_1 - \lambda_i \quad \text{for} \quad j = 1, \ldots, n. \tag{12}
$$

Since

$$
\sum_{j=1}^{n} x_{1j}^* = r_+ \quad \text{and} \quad \sum_{j=1}^{n} x_{ij}^* \geq r_-,
$$

there exists j such that

$$\frac{x_{ij}^*}{x_{1j}^*} \geq \frac{r_-}{r_+}.$$

We apply (12) with that index j. We have

$$\lambda_1 - \lambda_i = \ln \frac{(x_{1j}^* + 1)x_{ij}^*}{(x_{ij}^* + 1)x_{1j}^*}. \tag{13}$$

Now, the minimum value of

$$\frac{(a+1)b}{(b+1)a} \quad \text{where} \quad a \geq b \geq \tau a \quad \text{and} \quad a \geq \sigma$$

is attained at $a = \sigma$ and $b = \tau\sigma$ and equal to

$$\frac{\tau\sigma + \tau}{\tau\sigma + 1}.$$

In our case (13),

$$a = x_{1j}^*, \quad b = x_{ij}^*, \quad \sigma = \frac{c_-}{m}, \quad \tau = \frac{r_-}{r_+}, \quad \text{and} \quad \frac{\tau\sigma + \tau}{\tau\sigma + 1} = \frac{r_- c_- + mr_-}{r_- c_- + mr_+}.$$

Hence

$$\lambda_1 - \lambda_i \geq \ln \frac{r_- c_- + mr_-}{r_- c_- + mr_+}.$$

Therefore, for every j,

$$\ln\left(\frac{x_{ij}^* + 1}{x_{ij}^*}\right) = \ln\left(\frac{x_{1j}^* + 1}{x_{1j}^*}\right) - (\lambda_1 - \lambda_i)$$

$$\leq \ln\left(\frac{x_{1j}^* + 1}{x_{1j}^*}\right) - \ln \frac{r_- c_- + mr_-}{r_- c_- + mr_+}$$

$$\leq \ln \frac{c_- + m}{c_-} - \ln \frac{r_- c_- + mr_-}{r_- c_- + mr_+}.$$

Hence

$$\frac{x_{ij}^* + 1}{x_{ij}^*} \leq \frac{r_- c_- + r_+ m}{r_- c_-} \quad \text{for} \quad j = 1, \ldots, n$$

and

$$x_{ij}^* \geq \frac{r_- c_-}{r_+ m}$$

as desired. The second inequality in Part (1) is proved similarly.

To prove Part (2), we use an approach similar to that for Part (1), as well as its inequality. Let i_0 be the row such that $r_{i_0} = r_-$. Without loss of generality, we assume that $i_0 = 1$ and hence

$$x_{ij}^* \geq x_{1j}^* \quad \text{for} \quad j = 1, \ldots, n.$$

Thus we have

$$x_{1j}^* \leq \frac{c_j}{m} \leq \frac{c_+}{m} \quad \text{for} \quad j = 1, \ldots, n.$$

Next, we compare the entries of the i-th row of X^* and the entries of the first row using (12).
  Since

$$\sum_{j=1}^n x_{ij}^* \leq r_+ \quad \text{and} \quad \sum_{j=1}^n x_{1j}^* = r_-$$

there is j such that

$$\frac{x_{ij}^*}{x_{1j}^*} \leq \frac{r_+}{r_-}.$$

We apply (13) with that index j. The maximum value of

$$\frac{(a+1)b}{(b+1)a} \quad \text{where} \quad a \leq b \leq \tau a \quad \text{and} \quad a \geq \sigma$$

is attained at $a = \sigma, b = \tau\sigma$ and is equal to

$$\frac{\tau\sigma + \tau}{\tau\sigma + 1}.$$

In our case of (13),

$$a = x_{1j}^*, \quad b = x_{ij}^*, \quad \tau = \frac{r_+}{r_-}, \quad \sigma = \frac{r_- c_-}{r_+ m}, \quad \text{and} \quad \frac{\tau\sigma + \tau}{\tau\sigma + 1} = \frac{r_- c_- + mr_+}{r_- c_- + mr_-}$$

where the expression for σ follows by Part (1). Hence

$$\lambda_1 - \lambda_i \leq \ln \frac{r_- c_- + mr_+}{r_- c_- + mr_-}$$

and for all j we have

$$\ln \left(\frac{x_{ij}^* + 1}{x_{ij}^*} \right) = \ln \left(\frac{x_{1j}^* + 1}{x_{1j}^*} \right) - (\lambda_1 - \lambda_i)$$

$$\geq \ln \left(\frac{x_{1j}^* + 1}{x_{1j}^*} \right) - \ln \frac{r_- c_- + mr_+}{r_- c_- + mr_-}$$

$$\geq \ln \frac{c_+ + m}{c_+} - \ln \frac{r_- c_- + mr_+}{r_- c_- + mr_-}.$$

Hence

$$\frac{x_{ij}^* + 1}{x_{ij}^*} \geq \frac{(c_+ + m)(r_- c_- + mr_-)}{c_+(r_- c_- + mr_+)} \quad \text{for} \quad j = 1, \ldots, n$$

and the proof follows.   ∎

7. PROOF OF LEMMA 5.1

We will use the following bounds for the permanent.

7.1. The van der Waerden Bound

Let $B = (b_{ij})$ be an $N \times N$ doubly stochastic matrix, that is,

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

and

$$b_{ij} \geq 0 \quad \text{for} \quad i, j = 1, \ldots, N.$$

Then

$$\text{per } B \geq \frac{N!}{N^N}.$$

This is the famous van der Waerden bound proved by Falikman [21] and Egorychev [20], see also Chapter 12 of [34] and [28].

7.2. The Continuous Version of the Bregman-Minc Bound

Let $B = (b_{ij})$ be an $N \times N$ matrix such that

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

and

$$b_{ij} \geq 0 \quad i, j = 1, \ldots, N.$$

Furthermore, let

$$z_i = \max_{j=1,\ldots,N} b_{ij} > 0 \quad \text{for} \quad i = 1, \ldots, N.$$

Then

$$\text{per } B \leq \prod_{i=1}^{N} z_i \Gamma^{z_i} \left(\frac{1 + z_i}{z_i} \right).$$

This bound was obtained by Soules [39].

If $z_i = 1/r_i$ for integers r_i, the bound transforms into

$$\text{per } B \leq \prod_{i=1}^{N} \frac{(r_i!)^{1/r_i}}{r_i},$$

which can be easily deduced from the Minc conjecture proved by Bregman, see [11].

Now we are ready to prove Lemma 5.1.

Proof of Lemma 5.1.   The lower bound is the van der Waerden bound.

To prove the upper bound, define

$$f(\xi) = \xi \ln \Gamma \left(\frac{1+\xi}{\xi} \right) + \ln \xi \quad \text{for} \quad 0 < \xi \leq 1.$$

Then f is a concave function and by the Bregman-Minc bound, we have

$$\ln \text{per } B \leq \sum_{i=1}^{N} f(z_i).$$

The function

$$F(x) = \sum_{i=1}^{N} f(\xi_i) \quad \text{for} \quad x = (\xi_1, \ldots, \xi_N)$$

is concave on the simplex defined by the equation $\xi_1 + \cdots + \xi_N = \tau$ and inequalities $\xi_i \geq 0$ for $i = 1, \ldots, N$. It is also symmetric under permutations of $\xi_1, \ldots, \xi_N$. Hence the maximum of F is attained at

$$\xi_1 = \cdots = \xi_N = \tau/N,$$

and so

$$\ln \text{per } B \leq Nf \left(\frac{\tau}{N} \right).$$

Thus

$$\text{per } B \leq \left(\frac{\tau}{N} \right)^N \left(\Gamma(1 + \tau^{-1}N) \right)^\tau$$

and the rest follows by Stirling's formula. ∎

8. PROOF OF THEOREM 5.2

We begin our proof by restating a theorem of Bregman [11] in a slightly more general form.

Theorem 8.1.   *Let $Y = (y_{ij})$ be the positive $m \times n$ matrix that is the scaling of a positive $m \times n$ matrix $X = (x_{ij})$ to have margins (R, C). Then*

$$\sum_{ij} y_{ij} (\ln y_{ij} - \ln x_{ij}) \leq \sum_{ij} z_{ij} (\ln z_{ij} - \ln x_{ij})$$

for every matrix $Z \in \mathcal{P}(R, C)$, where $\mathcal{P}(R, C)$ is the transportation polytope of $m \times n$ non-negative matrices with row sums R and column sums C.

Proof.   The function

$$f(Z) = \sum_{ij} z_{ij}(\ln z_{ij} - \ln x_{ij})$$

is strictly convex on $\mathcal{P}(R, C)$ and hence attains its unique minimum $Y' = (y'_{ij})$ on $\mathcal{P}(R, C)$. As in the proof of Theorem 3.3 (see Section 6), we can show that Y' is strictly positive, that is, Y' lies in the relative interior of $\mathcal{P}(R, C)$. Writing the Lagrange multiplier conditions, we obtain

$$\ln y'_{ij} - \ln x_{ij} = \xi_i + \eta_j$$

for some $\xi_1, \ldots, \xi_m$ and $\eta_1, \ldots, \eta_n$. Letting $\lambda_i = e^{\xi_i}$ and $\mu_j = e^{\eta_j}$ we obtain

$$y'_{ij} = \lambda_i \mu_j x_{ij} \quad \text{for all} \quad i, j,$$

so in fact $Y' = Y$ as desired.   ∎

Next, we prove a lemma that extends a result of Linial, Samorodnitsky, and Wigderson [32].

Lemma 8.2.   *Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive vectors such that*

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

Let $X = (x_{ij})$ be an $m \times n$ positive matrix such that

$$\sum_{ij} x_{ij} = N$$

and let $Y = (y_{ij})$ be the scaling of X to have row sums R and column sums C. Then

$$\sum_{ij} r_i c_j \ln y_{ij} \geq \sum_{ij} r_i c_j \ln x_{ij}.$$

Proof.   Since Y is the limit of the sequence of matrices obtained from X by repeated alternate scaling of the rows to have row sums $r_1, \ldots, r_m$ and of the columns to have column sums $c_1, \ldots, c_n$, cf., for example, Chapter 6 of [4], it suffices to show that when the rows (columns) are scaled, the corresponding weighted sums of the logarithms of the entries of the matrix can only increase.

To this end, let $X = (x_{ij})$ be a positive $m \times n$ matrix with the row sums $\sigma_1, \ldots, \sigma_m$ such that

$$\sum_{i=1}^{m} \sigma_i = N$$

and let $Y = (y_{ij})$ be the matrix obtained from Y by scaling the rows to have sums $r_1, \ldots, r_m$. Hence,

$$y_{ij} = r_i x_{ij}/\sigma_i \quad \text{for all} \quad i, j.$$

Thus

$$\sum_{ij} r_i c_j (\ln y_{ij} - \ln x_{ij}) = \sum_{j=1}^{n} c_j \left(\sum_{i=1}^{m} (r_i \ln r_i - r_i \ln \sigma_i) \right) \geq 0,$$

since the maximum of the function

$$\sum_{i=1}^{m} r_i \ln \xi_i$$

on the simplex

$$\left\{ \sum_{i=1}^{m} \xi_i = N \quad \text{and} \quad \xi_i \geq 0 \quad \text{for} \quad i = 1, \ldots, m \right\}$$

is attained at $\xi_i = r_i$.

The scaling of columns is treated similarly. ∎

Proof of Theorem 5.2. Without loss of generality, we assume that $p = q = 1$.

Define an $m \times n$ matrix $U = (u_{ij})$ by

$$u_{ij} = \frac{r_i c_j x_{ij}}{T} \quad \text{for} \quad T = \frac{1}{N} \sum_{ij} r_i c_j x_{ij}. \tag{14}$$

We note that the scalings of U and X to margins (R, C) coincide and that

$$\sum_{ij} u_{ij} = N.$$

By Theorem 8.1, the matrix Y minimizes

$$\sum_{ij} z_{ij} (\ln z_{ij} - \ln u_{ij}),$$

over the set $\mathcal{P}(R, C)$ of $m \times n$ non-negative matrices Z with row sums R and the column sums C.

For a real t, let us define the matrix $Y(t) = (y_{ij}(t))$ by

$$y_{ij}(t) = \begin{cases} y_{ij} + t & \text{if} \quad i = j = 1 \\ y_{ij} - \dfrac{c_j}{N - c_1} t & \text{if} \quad i = 1, j \neq 1 \\ y_{ij} - \dfrac{r_i}{N - r_1} t & \text{if} \quad i \neq 1, j = 1 \\ y_{ij} + \dfrac{r_i c_j}{(N - r_1)(N - c_1)} t & \text{if} \quad i \neq 1, j \neq 1. \end{cases}$$

Then $Y(0) = Y$ and $Y(t) \in \mathcal{P}(R, C)$ for all t sufficiently close to 0. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} f(Y(t)) \Big|_{t=0} = 0,$$

where

$$f(Z) = \sum_{ij} z_{ij}(\ln z_{ij} - \ln u_{ij}).$$

Therefore,

$$\ln y_{11} - \ln u_{11} + 1 - \frac{1}{N - c_1} \sum_{j \neq 1} c_j(\ln y_{1j} - \ln u_{1j} + 1)$$

$$- \frac{1}{N - r_1} \sum_{i \neq 1} r_i(\ln y_{i1} - \ln u_{i1} + 1)$$

$$+ \frac{1}{(N - r_1)(N - c_1)} \sum_{i,j \neq 1} r_i c_j(\ln y_{ij} - \ln u_{ij} + 1)$$

$$= 0.$$

Rearranging the summands,

$$\frac{N^2}{(N - r_1)(N - c_1)}(\ln y_{11} - \ln u_{11}) - \frac{N}{(N - r_1)(N - c_1)} \sum_{j=1}^{n} c_j(\ln y_{1j} - \ln u_{1j})$$

$$- \frac{N}{(N - r_1)(N - c_1)} \sum_{i=1}^{m} r_i(\ln y_{i1} - \ln u_{i1}) + \frac{1}{(N - r_1)(N - c_1)} \sum_{ij} r_i c_j(\ln y_{ij} - \ln u_{ij})$$

$$= 0.$$

On the other hand, by Lemma 8.2,

$$\sum_{ij} r_i c_j(\ln y_{ij} - \ln u_{ij}) \geq 0,$$

so we must have

$$N^2(\ln y_{11} - \ln u_{11}) - N \sum_{j=1}^{n} c_j(\ln y_{1j} - \ln u_{1j}) - N \sum_{i=1}^{m} r_i(\ln y_{i1} - \ln u_{i1}) \leq 0.$$

In other words,

$$\ln y_{11} \leq \ln u_{11} + \frac{1}{N} \sum_{j=1}^{n} c_j(\ln y_{1j} - \ln u_{1j}) + \frac{1}{N} \sum_{i=1}^{m} r_i(\ln y_{i1} - \ln u_{i1}).$$

Since

$$\sum_{j=1}^{n} y_{1j} = r_1,$$

we have

$$\sum_{j=1}^{n} c_j \ln y_{1j} \le \sum_{j=1}^{n} c_j \ln \left(\frac{c_j r_1}{N} \right),$$

cf. the proof of Lemma 8.2. Similarly, since

$$\sum_{i=1}^{m} y_{i1} = c_1,$$

we have

$$\sum_{i=1}^{m} r_i \ln y_{i1} \le \sum_{i=1}^{m} r_i \ln \left(\frac{r_i c_1}{N} \right).$$

Substituting (14) for U, we obtain

$$\ln y_{11} \le \ln x_{11} + \ln(r_1 c_1) - \ln T + \frac{1}{N} \sum_{j=1}^{n} c_j \ln \frac{T}{N x_{1j}} + \frac{1}{N} \sum_{i=1}^{m} r_i \ln \frac{T}{N x_{i1}},$$

and the proof follows.   ∎

9. PROOF OF THEOREM 5.3

Fix margins (R, C), let $\psi = \psi_{R,C}$ be the density of Section 2.2, and let $X = (x_{ij})$ be the random matrix distributed in accordance with the density ψ. We will need a lemma that connects linear functionals of X with the weighted sums $T(R, C; W)$ of Section 4.1.

Lemma 9.1.   *Let $\lambda_{ij} < 1$ be real numbers.*

1. *Let $W = (w_{ij})$ be the $m \times n$ matrix of weights given by*

$$w_{ij} = (1 - \lambda_{ij})^{-1} \quad for\ all \quad i, j.$$

   Then

$$\mathbf{E} \exp \left\{ \sum_{ij} \lambda_{ij} x_{ij} \right\} = \frac{T(R, C; W)}{\#(R, C)} \prod_{ij} w_{ij};$$

2. *We have*

$$\mathbf{E} \prod_{ij} x_{ij}^{-\lambda_{ij}} = \frac{1}{\#(R, C)} \sum_{D=(d_{ij})} \prod_{ij} \frac{\Gamma(d_{ij} - \lambda_{ij} + 1)}{\Gamma(d_{ij} + 1)},$$

   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.

Proof. Let us prove Part (1). We have

$$\mathbf{E}\exp\left\{\sum_{ij}\lambda_{ij}x_{ij}\right\} = \frac{1}{\#(R,C)}\int_{\mathbb{R}_+^{mn}}\exp\left\{-\sum_{ij}(1-\lambda_{ij})x_{ij}\right\}\sum_{D=(d_{ij})}\prod_{ij}\frac{x_{ij}^{d_{ij}}}{d_{ij}!}\,dX$$

$$= \frac{1}{\#(R,C)}\int_{\mathbb{R}_+^{mn}}\exp\left\{-\sum_{ij}x_{ij}\right\}\sum_{D=(d_{ij})}\prod_{ij}\frac{w_{ij}^{d_{ij}}x_{ij}^{d_{ij}}}{d_{ij}!}\prod_{ij}w_{ij}\,dX$$

$$= \frac{T(R,C;W)}{\#(R,C)}\prod_{ij}w_{ij},$$

as desired.

Since

$$\psi(X)\prod_{ij}x_{ij}^{-\lambda_{ij}} = \frac{1}{\#(R,C)}\sum_{D=(d_{ij})}\prod_{ij}\frac{x_{ij}^{d_{ij}-\lambda_{ij}}}{d_{ij}!}e^{-x_{ij}},$$

the proof of Part (2) follows. ∎

To prove Theorem 5.3 we need only Part (1) of the lemma, while Part (2) will be used later in the proof of Theorem 3.2.

Proof of Theorem 5.3. We use the Laplace transform method, see, for example, Appendix A of [1]. We have

$$\mathbf{P}\left\{\sum_{(i,j)\in S}x_{ij}\geq t\right\} = \mathbf{P}\left\{\exp\left\{\frac{1}{2\lambda+2}\sum_{(i,j)\in S}x_{ij}\right\}\geq\exp\left\{\frac{t}{2\lambda+2}\right\}\right\}$$

$$\leq \exp\left\{-\frac{t}{2\lambda+2}\right\}\mathbf{E}\exp\left\{\frac{1}{2\lambda+2}\sum_{(i,j)\in S}x_{ij}\right\},$$

by the Markov inequality.

By Part (1) of Lemma 9.1,

$$\mathbf{E}\exp\left\{\frac{1}{2\lambda+2}\sum_{(i,j)\in S}x_{ij}\right\} = \frac{T(R,C;W)}{\#(R,C)}\left(\frac{2\lambda+2}{2\lambda+1}\right)^{\#S},$$

where

$$w_{ij} = \begin{cases}(2\lambda+2)/(2\lambda+1) & \text{if } (i,j)\in S \\ 1 & \text{if } (i,j)\notin S.\end{cases}$$

Clearly,

$$\left(\frac{2\lambda+2}{2\lambda+1}\right)^{\#S}\leq 2^{\#S}.$$

To bound the ratio of $T(R,C;W)$ and $\#(R,C)$, we use Theorems 4.1 and 4.2.

Let $0 < x_1, \ldots, x_m; y_1, \ldots, y_n < 1$ be numbers such that

$$\rho(R, C) = \left(\prod_{i=1}^{m} x_i^{-r_i} \right) \left(\prod_{j=1}^{n} y_j^{-c_j} \right) \left(\prod_{ij} \frac{1}{1 - x_i y_j} \right).$$

For the typical table $X^* = (x_{ij}^*)$ we have

$$x_{ij}^* = \frac{x_i y_j}{1 - x_i y_j} \leq \lambda \quad \text{for all} \quad i, j.$$

Therefore,

$$x_i y_j = \frac{x_{ij}^*}{1 + x_{ij}^*} \leq \frac{\lambda}{\lambda + 1} \quad \text{for all} \quad i, j$$

and

$$w_{ij} x_i y_j < 1 \quad \text{for all} \quad i, j.$$

Then we have

$$\rho(R, C; W) \leq \left(\prod_{i=1}^{m} x_i^{-r_i} \right) \left(\prod_{j=1}^{n} y_j^{-c_j} \right) \left(\prod_{ij} \frac{1}{1 - w_{ij} x_i y_j} \right)$$

and

$$\frac{\rho(R, C; W)}{\rho(R, C)} \leq \prod_{(i,j) \in S} \frac{1 - x_i y_j}{1 - w_{ij} x_i y_j} = \prod_{(i,j) \in S} \frac{1}{1 + (1 - w_{ij}) x_{ij}^*}.$$

Now

$$\frac{1}{1 + (1 - w_{ij}) x_{ij}^*} \leq \frac{2\lambda + 1}{\lambda + 1} \leq 2 \quad \text{for all} \quad (i, j) \in S$$

and hence

$$\frac{\rho(R, C; W)}{\rho(R, C)} \leq 2^{\#S}.$$

Since

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

the proof follows.   ∎

We will need the following corollary.

Corollary 9.2. *Suppose that $m \geq n$ and that the typical table $X^* = (x_{ij}^*)$ satisfies*

$$x_{ij}^* \leq \lambda \quad \text{for all} \quad i, j$$

and some $\lambda > 0$. Let $X = (x_{ij})$ be a random $m \times n$ matrix distributed in accordance with the density $\psi_{R,C}$, and let

$$u_i = \max_{j=1,\ldots,n} x_{ij}.$$

Then for some $\tau = \tau(\lambda) > 0$ we have

$$\mathbf{P} \left\{ \sum_{i=1}^{m} u_i \geq (\lambda + 1)\tau m \ln N \right\} \leq 4^{-m}.$$

Proof. We apply Theorem 5.3 to each of the n^m of subsets S having exactly one entry in each row and sum up the tail estimates. Hence we obtain

$$\mathbf{P} \left\{ \sum_{i=1}^{m} u_i \geq (\lambda + 1)\tau m \ln N \right\} \leq n^m N^{-\tau m/2} 4^m N^{\gamma(m+n)}$$

$$\leq 4^{-m}$$

for a sufficiently large absolute constant τ. ∎

We will also use an unconditional bound on the sum of all the entries of X.

Lemma 9.3. *Let $X = (x_{ij})$ be a random $m \times n$ matrix distributed in accordance with the density $\psi_{R,C}$. Then*

$$\mathbf{P} \left\{ \sum_{ij} x_{ij} \geq 2(N + mn) \right\} \leq \left(\frac{3}{4} \right)^{N+mn}.$$

Proof. As in the proof of Theorem 5.3, we have

$$\mathbf{P} \left\{ \sum_{ij} x_{ij} \geq 2(N + mn) \right\} = \mathbf{P} \left\{ \exp \left\{ \frac{1}{2} \sum_{ij} x_{ij} \right\} \geq \exp\{N + mn\} \right\}$$

$$\leq \exp\{-(N + mn)\} \mathbf{E} \exp \left\{ \frac{1}{2} \sum_{ij} x_{ij} \right\}$$

by Markov's inequality. By Lemma 9.1,

$$\mathbf{E} \exp \left\{ \frac{1}{2} \sum_{ij} x_{ij} \right\} = \frac{T(R, C; W)}{\#(R, C)} \prod_{ij} w_{ij} \quad \text{where}$$

$$w_{ij} = 2 \quad \text{for all} \quad i, j$$

$$= 2^{N+mn}$$

and the proof follows. ∎

10. PROOF OF THEOREM 3.1

We start with a technical result.

Lemma 10.1.   *Let (R, C) be upper α-smooth margins, so $r_i/N \leq \alpha/m$ and $c_j/N \leq \alpha/n$ for all i, j. Let $X = (x_{ij})$ be a random $m \times n$ matrix with density $\psi_{R,C}$ of Section 2.2. Then for any real τ*

$$\mathbf{P}\left\{\frac{1}{N}\sum_{j=1}^{n} c_j \ln x_{ij} \leq -\tau\right\} \leq 2^n \exp\left\{-\frac{n\tau}{2\alpha}\right\} \quad and$$

$$\mathbf{P}\left\{\frac{1}{N}\sum_{i=1}^{m} r_i \ln x_{ij} \leq -\tau\right\} \leq 2^m \exp\left\{-\frac{m\tau}{2\alpha}\right\}.$$

Proof.   Let us prove the first inequality. As in the proof of Theorem 5.3, we use the Laplace transform method. We have

$$\mathbf{P}\left\{\frac{1}{N}\sum_{j=1}^{n} c_j \ln x_{ij} \leq -\tau\right\} = \mathbf{P}\left\{-\frac{n}{2\alpha N}\sum_{j=1}^{n} c_j \ln x_{ij} \geq \frac{n\tau}{2\alpha}\right\}$$

$$\leq \exp\left\{-\frac{n\tau}{2\alpha}\right\}\mathbf{E}\,\exp\left\{-\frac{n}{2\alpha N}\sum_{j=1}^{n} c_j \ln x_{ij}\right\}$$

$$= \exp\left\{-\frac{n\tau}{2\alpha}\right\}\mathbf{E}\prod_{j=1}^{n} x_{ij}^{-\lambda_j} \quad \text{where} \quad \lambda_j = \frac{nc_j}{2\alpha N}.$$

Since

$$\lambda_j \leq \frac{1}{2},$$

by Part (2) of Lemma 9.1 we deduce that

$$\mathbf{E}\prod_{j=1}^{n} x_{ij}^{-\lambda_j} \leq \left(\Gamma\left(\frac{1}{2}\right)\right)^n \leq 2^n$$

(we observe that every term in the sum of Lemma 9.1 does not exceed $\Gamma^n(1/2)$). The proof of the second inequality is identical.   ∎

Proof of Theorem 3.1.   Without loss of generality, we assume that $m \geq n$. We recall that function $p(X)$ is computed as follows. Given a positive $m \times n$ matrix $X = (x_{ij})$, we compute the scaling $Y = (y_{ij})$ of X to have row sums R and the column sums C. Then we compute the $N \times N$ block matrix $B(X)$ consisting of mn blocks of sizes $r_i \times c_j$ with the entries in the (i, j)-th block equal to $y_{ij}/r_i c_j$. Thus $B(X)$ is a doubly stochastic matrix and

$$p(X) = \frac{N^N}{N!}\text{per}\, B(X),$$

cf. Section 2.

We are going to use Theorem 5.2 to bound the entries of Y.

By Lemma 9.3,

$$\mathbf{P}\left\{\sum_{ij} x_{ij} < 2(N+mn)\right\} \geq 1 - \left(\frac{3}{4}\right)^{N+mn}.$$

Since $N \leq s_0 mn$, $r_i/N \leq \alpha/m$, and $c_j/N \leq \alpha/n$ we conclude that for some $\kappa_1 = \kappa_1(\alpha, s_0) = 2\alpha^2(s_0 + 1)$ we have

$$\mathbf{P}\left\{\frac{1}{N^2}\sum_{ij} r_i c_j x_{ij} < \kappa_1\right\} \geq 1 - \left(\frac{3}{4}\right)^{N+mn}.$$

From Lemma 10.1, for a sufficiently large $\kappa_2 = \kappa_2(\alpha)$, we have

$$\mathbf{P}\left\{\frac{1}{N}\sum_{j=1}^{n} c_j \ln x_{pj} > -\kappa_2\right\} \geq 1 - 4^{-n} \quad \text{for all} \quad p = 1, \ldots, m \quad \text{and}$$

$$\mathbf{P}\left\{\frac{1}{N}\sum_{i=1}^{m} r_i \ln x_{iq} > -\kappa_2\right\} \geq 1 - 4^{-m} \quad \text{for} \quad q = 1, \ldots, n.$$

Therefore, by Theorem 5.2, we have for some $\kappa = \kappa(\alpha, s_0)$

$$\mathbf{P}\left\{y_{pq} \leq \frac{r_p c_q}{N}\kappa x_{pq} \quad \text{for all} \quad p, q\right\} \geq 1 - \left(\frac{3}{4}\right)^{N+nm} - m4^{-n} - n4^{-m}.$$

Now, B consists of mn blocks, where the (p, q)-th block filled by the entries $y_{pq}/r_p c_q$. Therefore the probability that for all $i, j = 1, \ldots N$ we have

$$b_{ij} \leq \frac{\kappa}{N}x_{pq} \quad \text{provided} \quad (i, j) \quad \text{lies in the } (p, q)\text{-th block of } B \tag{15}$$

is at least

$$1 - \left(\frac{3}{4}\right)^{N+nm} - m4^{-n} - n4^{-m}.$$

We now bound per $B(X)$ using Lemma 5.1 and Corollary 9.2.

Let

$$z_i = \max_{j=1,\ldots,N} b_{ij} \quad \text{for} \quad i = 1, \ldots N \quad \text{and let}$$

$$u_p = \max_{q=1,\ldots,m} x_{pq}.$$

Then, from (15) we have

$$\sum_{i=1}^{N} z_i \leq \frac{\kappa}{N}\sum_{p=1}^{m} r_p u_p \leq \frac{\alpha\kappa}{m}\sum_{p=1}^{m} u_p.$$

By Corollary 9.2, for some $\tau_1 = \tau_1(\alpha, s_0)$, we have

$$\mathbf{P}\left\{\sum_{p=1}^{m} u_m \leq \tau_1 m \ln N\right\} \geq 1 - 4^{-m}.$$

Thus for some $\tau = \tau(\alpha, s_0)$ we have

$$\mathbf{P}\left\{\sum_{i=1}^{N} z_i \leq \tau \ln N\right\} \geq 1 - \left(\frac{3}{4}\right)^{N+mn} - m4^{-n} - n4^{-m} - 4^{-m}$$

and the proof follows by Lemma 5.1. ∎

The rest of the paper deals with the proof of Theorem 3.2. This requires sharpening of the estimates of Lemma 10.1. Roughly, we need to prove that with overwhelming probability

$$\frac{1}{N}\sum_{j=1}^{n} c_j \ln x_{ij} \geq -\tau + \ln s \quad \text{and}$$

$$\frac{1}{N}\sum_{i=1}^{m} r_i \ln x_{ij} \geq -\tau + \ln s$$

for some constant $\tau = \tau(\alpha, \beta)$, where $s = N/mn$ is the average entry of the table.

11. AN ESTIMATE OF A SUM OVER TABLES

To sharpen the estimates of Lemma 10.1 we need a more careful estimate of the sum in Part (2) of Lemma 9.1. In this section, we prove the following technical result.

Proposition 11.1. *Suppose that (R, C) are lower β-smooth and upper α-smooth margins and that*

$$s = N/mn \geq 1.$$

Let $\lambda_1, \ldots, \lambda_m \leq 1/2$ be numbers and let $l = \lambda_1 + \cdots + \lambda_m$. Then, for $k < n$ we have

$$\frac{1}{\#(R, C)}\sum_{D=(d_{ij})} \prod_{\substack{1 \leq i \leq m \\ 1 \leq j \leq k}} \frac{\Gamma(d_{ij} - \lambda_i + 1)}{\Gamma(d_{ij} + 1)} \leq \delta^{km} N^{\gamma(m+n)} s^{-kl},$$

where the sum is taken over all non-negative integer matrices D with row sums R and column sums C, $\delta = \delta(\alpha, \beta) > 0$ and γ is the absolute constant of Theorem 4.1.

We start with computing a simplified version of this sum in a closed form.

Definition 11.2. *Let us fix positive integers c and m. The integer simplex $\Upsilon(m, c)$ is the set of all non-negative integer vectors $a = (d_1, \ldots, d_m)$ such that $d_1 + \cdots + d_m = c$.*
Clearly,

$$\#\Upsilon(m, c) = \binom{m + c - 1}{m - 1}.$$

A sum over $\Upsilon(m, c)$ similar to that of Proposition 11.1 can be computed in a closed form.

Lemma 11.3. *Let $\lambda_i < 1$, $i = 1, \ldots, m$, be numbers and let $l = \lambda_1 + \cdots + \lambda_m$. Then*

$$\frac{1}{\#\Upsilon(m,c)} \sum_{\substack{d_1,\ldots,d_m \geq 0 \\ d_1+\ldots+d_m=c}} \prod_{i=1}^{m} \frac{\Gamma(d_i - \lambda_i + 1)}{\Gamma(d_i + 1)} = \frac{\Gamma(c+m-l)\Gamma(m)}{\Gamma(c+m)\Gamma(m-l)} \prod_{i=1}^{m} \Gamma(1 - \lambda_i).$$

Proof. Let us define a function h_c on the positive orthant $\mathbb{R}_+^m$ by the formula

$$h_c(x) = \frac{(m-1)!}{(m+c-1)!} \left(\sum_{i=1}^{m} \xi_i \right)^c \exp\left\{ -\sum_{i=1}^{m} \xi_i \right\} \quad \text{for} \quad x = (\xi_1, \ldots, \xi_m) \in \mathbb{R}_+^m.$$

Since

$$\left(\sum_{i=1}^{m} \xi_i \right)^c = \sum_{\substack{d_1,\ldots,d_m \geq 0 \\ d_1+\cdots+d_m=c}} \frac{c!}{d_1! \cdots d_m!} \xi_1^{d_1} \cdots \xi_m^{d_m},$$

We can rewrite

$$h_c(x) = \binom{m+c-1}{m-1}^{-1} \sum_{\substack{d_1,\ldots,d_m \geq 0 \\ d_1+\cdots+d_m=c}} \prod_{i=1}^{m} \frac{\xi_i^{d_i}}{d_i!} e^{-\xi_i}.$$

Therefore,

$$\frac{1}{\#\Upsilon(m,c)} \sum_{\substack{d_1,\ldots,d_m \geq 0 \\ d_1+\cdots+d_m=c}} \prod_{i=1}^{m} \frac{\Gamma(d_i - \lambda_i + 1)}{\Gamma(d_i + 1)} = \int_{\mathbb{R}_+^m} h_c(x) \prod_{i=1}^{m} \xi_i^{-\lambda_i} \, dx.$$

Let $Q \subset \mathbb{R}_+^m$ be the simplex $\xi_1 + \cdots + \xi_m = 1$ with the Lebesgue measure dx normalized to the probability measure. Since the function

$$\left(\sum_{i=1}^{m} \xi_i \right)^c \prod_{i=1}^{m} \xi_i^{-\lambda_i}$$

is positive homogeneous of degree $c - l$, we can write

$$\int_{\mathbb{R}_+^m} h_c(x) \prod_{i=1}^{m} \xi_i^{-\lambda_i} \, dx = \frac{\Gamma(c+m-l)}{\Gamma(m)} \int_Q h_c(x) \prod_{i=1}^{m} \xi_i^{-\lambda_i} \, dx \tag{16}$$

On the other hand,

$$h_c(x) = \frac{\Gamma(m)}{\Gamma(c+m)} h_0(x) \quad \text{for} \quad x \in Q. \tag{17}$$

Using (16) with $c = 0$, we deduce that

$$\int_Q h_0(x) \prod_{i=1}^m \xi_i^{-\lambda_i} \, dx = \frac{\Gamma(m)}{\Gamma(m-l)} \int_{\mathbb{R}_+^m} h_0(x) \prod_{i=1}^m \xi_i^{-\lambda_i} \, dx$$

$$= \frac{\Gamma(m)}{\Gamma(m-l)} \prod_{i=1}^m \int_0^{+\infty} \xi_i^{-\lambda_i} e^{-\xi_i} \, d\xi_i$$

$$= \frac{\Gamma(m)}{\Gamma(m-l)} \prod_{i=1}^m \Gamma(1-\lambda_i).$$

Now, from (16) and (17), we have

$$\int_{\mathbb{R}_+^m} h_c(x) \prod_{i=1}^m \xi_i^{-\lambda_i} \, dx = \frac{\Gamma(c+m-l)\Gamma(m)}{\Gamma(c+m)\Gamma(m-l)} \prod_{i=1}^m \Gamma(1-\lambda_i),$$

as desired. ∎

We need an estimate.

Corollary 11.4. *Suppose that $\lambda_i < 1/2$ for $i = 1, \dots, m$ and $c \geq \beta m$ for some $\beta > 0$. Then*

$$\frac{1}{\#\Upsilon(m,c)} \sum_{\substack{d_1,\dots,d_m \geq 0 \\ d_1+\cdots+d_m=c}} \prod_{i=1}^m \frac{\Gamma(d_i - \lambda_i + 1)}{\Gamma(d_i + 1)} \leq \left(\frac{m}{c}\right)^l \delta^m$$

for some constant $\delta = \delta(\beta) > 0$, where $l = \lambda_1 + \cdots + \lambda_m$.

Proof. The proof follows from Lemma 11.3. ∎

Fix margins $R = (r_1, \dots, r_m)$ and $C = (c_1, \dots, c_n)$ and a number $k \leq n$. Pick, uniformly at random, a contingency table $D = (d_{ij})$ with margins (R, C) and consider its submatrix Z consisting of the first k columns. Hence Z is an $m \times k$ non-negative integer matrix with the column sums $c_1, \dots, c_k$. We interpret Z as a point in the product

$$\Upsilon = \Upsilon(m, c_1) \times \cdots \times \Upsilon(m, c_k)$$

of integer simplices. This process induces a certain distribution on the set Υ of non-negative integer $m \times k$ matrices with the column sums $c_1, \dots, c_k$. We want to compare this distribution with the uniform distribution. Lemma 11.5 below says that the probability to get any particular matrix $Z \in \Upsilon$ cannot exceed the uniform probability by much if the margins (R, C) are smooth.

Once we fix the $m \times k$ submatrix Z consisting of the first k columns of a table with margins (R, C), the complementary $m \times (n-k)$ table has row sums $R' = R - R(Z)$, where $R(Z)$ is the vector of row sums of Z, and column sums $\bar{C} = (c_{k+1}, \dots, c_n)$, the truncation of C. Hence the probability of obtaining a particular $Z \in \Upsilon$ is

$$\frac{\#(R', \bar{C})}{\#(R, C)},$$

where the ratio is declared to be 0 if R' is not non-negative.

We prove the following estimate.

Lemma 11.5. *Consider margins (R, C) satisfying the constraints of Proposition 11.1. Fix $k \leq n$ and let Υ be the set of all $m \times k$ non-negative integer matrices with the column sums $c_1, \ldots, c_k$.*

Let $\bar{C} = (c_{k+1}, \ldots, c_n)$, choose $Z \in \Upsilon$ and set $R' = R - R(Z)$, where $R(Z)$ is the vector of the row sums of Z. Then

$$\frac{\#(R', \bar{C})}{\#(R, C)} \leq \frac{\delta^{km} N^{\gamma(m+n)}}{\#\Upsilon}$$

for some constant $\delta = (\alpha, \beta) > 0$, where $\gamma > 0$ is an absolute constant from Theorem 4.1.

Proof. Let $\rho(R, C)$ be the quantity of Theorem 4.1. Here we agree that $\rho(R', \bar{C}) = 0$ if R' has negative components and that "max" and "min" are replaced by "sup" and "inf," respectively, if R' is non-negative but has 0 components.

Let $0 < x_1, \ldots, x_m < 1$ and $0 < y_1, \ldots, y_n < 1$ be an optimal point in Theorem 4.1, so

$$\rho(R, C) = \prod_{i=1}^{m} x_i^{-r_i} \prod_{j=1}^{n} y_j^{-c_j} \prod_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \frac{1}{1 - x_i y_j}.$$

Then

$$\rho(R', \bar{C}) \leq \prod_{i=1}^{m} x_i^{-r'_i} \prod_{j=k+1}^{n} y_j^{-c_j} \prod_{\substack{1 \leq i \leq m \\ k+1 \leq j \leq n}} \frac{1}{1 - x_i y_j}$$

$$\leq \prod_{i=1}^{m} x_i^{-r_i} \prod_{j=1}^{n} y_j^{-c_j} \prod_{\substack{1 \leq i \leq m \\ k+1 \leq j \leq n}} \frac{1}{1 - x_i y_j}$$

and hence

$$\frac{\rho(R', \bar{C})}{\rho(R, C)} \leq \prod_{\substack{1 \leq i \leq m \\ 1 \leq j \leq k}} (1 - x_i y_j).$$

Now, by Part (1) of Theorem 3.3, the typical table $X^* = (x_{ij}^*)$ satisfies

$$x_{ij}^* = \frac{x_i y_j}{1 - x_i y_j} \geq \delta_1 s \quad \text{for all} \quad i, j,$$

and for some $\delta_1 = \delta_1(\alpha, \beta)$. This implies that

$$1 - x_i y_j = \frac{1}{1 + x_{ij}^*} \leq \frac{1}{1 + \delta_1 s} \quad \text{for all} \quad i, j.$$

Summarizing,

$$\frac{\rho(R', \bar{C})}{\rho(R, C)} \leq \left(\frac{1}{1 + \delta_1 s} \right)^{km}.$$

Now,

$$\#\Upsilon = \prod_{j=1}^{k} \binom{c_j + m - 1}{m - 1} \le \prod_{j=1}^{k} \binom{c_j + m}{m}$$

$$\le \prod_{j=1}^{k} \left(\frac{c_j + m}{c_j} \right)^{c_j} \left(\frac{c_j + m}{m} \right)^{m}.$$

We have

$$\left(\frac{c_j + m}{c_j} \right)^{c_j} \le e^m.$$

Furthermore, since $c_j \le \alpha sm$, we have

$$\left(\frac{c_j + m}{m} \right)^{m} \le (1 + \alpha s)^m$$

and

$$\#\Upsilon \, \frac{\rho(R', \bar{C})}{\rho(R, C)} \le e^{km} \left(\frac{1 + \alpha s}{1 + \delta_1 s} \right)^{km} \le \delta^{km}.$$

Since by Theorem 4.1 we have

$$\#(R, C) \ge N^{-\gamma(m+n)} \rho(R, C) \quad \text{and} \quad \#(R', \bar{C}) \le \rho(R', \bar{C}),$$

the proof follows. ∎

Proof of Proposition 11.1. Let $\Upsilon(m, c_j)$ be the integer simplex of non-negative integer vectors summing up to c_j and let

$$\Upsilon = \Upsilon(m, c_1) \times \cdots \times \Upsilon(m, c_k).$$

Using Lemma 11.5, we bound

$$\frac{1}{\#(R, C)} \sum_{\substack{D = (d_{ij}) \\ 1 \le i \le m \\ 1 \le j \le k}} \prod \frac{\Gamma(d_{ij} - \lambda_i + 1)}{\Gamma(d_{ij} + 1)} = \sum_{\substack{Z = (z_{ij}) \\ Z \in \Upsilon}} \frac{\#(R - R(Z), \bar{C})}{\#(R, C)} \prod_{\substack{1 \le i \le m \\ 1 \le j \le k}} \frac{\Gamma(z_{ij} - \lambda_i + 1)}{\Gamma(z_{ij} + 1)}$$

$$\le \frac{\delta_1^{km} N^{\gamma(m+n)}}{\#\Upsilon} \sum_{\substack{Z = (z_{ij}) \\ Z \in \Upsilon}} \prod_{\substack{1 \le i \le m \\ 1 \le j \le k}} \frac{\Gamma(z_{ij} - \lambda_i + 1)}{\Gamma(z_{ij} + 1)}$$

for some $\delta_1 = \delta(\alpha, \beta)$. The sum

$$\frac{1}{\#\Upsilon} \sum_{\substack{Z = (z_{ij}) \\ Z \in \Upsilon}} \prod_{\substack{1 \le i \le m \\ 1 \le j \le k}} \frac{\Gamma(z_{ij} - \lambda_i + 1)}{\Gamma(z_{ij} + 1)}$$

is just the product of k sums of the type

$$\frac{1}{\Upsilon(m, c_j)} \sum_{\substack{d_1, \ldots, d_m \ge 0 \\ d_1 + \cdots + d_m = c_j}} \prod_{i=1}^{m} \frac{\Gamma(d_i - \lambda_i + 1)}{\Gamma(d_i + 1)} \le \left(\frac{m}{c_j} \right)^{l} \delta_2^m$$

by Corollary 11.4, for some $\delta_2 = \delta(\alpha, \beta)$. The proof now follows. ∎

12. PROOF OF THEOREM 3.2

Fix margins (R, C) and let $X = (x_{ij})$ be the $m \times n$ random matrix with density $\psi = \psi_{R,C}$ of Section 2.2. Define random variables

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

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

Lemma 12.1. *Let (R, C) be lower β-smooth upper α-smooth margins such that $s = N/mn \geq 1$.*

Choose a subset $J \subset \{1, \dots, n\}$ of indices, $\#J = k$. Then for all $t > 0$ we have

$$\mathbf{P} \left\{ \frac{1}{k} \sum_{j \in J} v_j \leq -t + \ln s \right\} \leq \exp\left\{ -\frac{tkm}{2\alpha} \right\} \delta^{km} N^{\gamma(m+n)},$$

Similarly, for a subset $I \subset \{1, \dots, m\}$ of indices, $\#I = k$, we have

$$\mathbf{P} \left\{ \frac{1}{k} \sum_{i \in I} h_i \leq -t + \ln s \right\} \leq \exp\left\{ -\frac{tkn}{2\alpha} \right\} \delta^{kn} N^{\gamma(m+n)}.$$

for some number $\delta = \delta(\alpha, \beta) > 0$ and the absolute constant $\gamma > 0$ of Theorem 4.1.

Proof. Without loss of generality, it suffices to prove only the first bound and only in the case of $J = \{1, \dots, k\}$.

We use the Laplace transform method. We have

$$\mathbf{P} \left\{ \frac{1}{k} \sum_{j=1}^{k} v_j \leq -t + \ln s \right\} = \mathbf{P} \left\{ -\frac{m}{2\alpha} \sum_{j=1}^{k} v_j \geq \frac{tkm}{2\alpha} - \frac{km \ln s}{2\alpha} \right\}$$

$$= \mathbf{P} \left\{ \exp\left\{ -\frac{m}{2\alpha} \sum_{j=1}^{k} v_j \right\} \geq s^{-km/2\alpha} \cdot \exp\left\{ \frac{tkm}{2\alpha} \right\} \right\}$$

$$\leq s^{km/2\alpha} \exp\left\{ -\frac{tkm}{2\alpha} \right\} \cdot \mathbf{E} \exp\left\{ -\frac{m}{2\alpha} \sum_{j=1}^{k} v_j \right\}.$$

Let

$$\lambda_i = \frac{mr_i}{2\alpha N} \leq \frac{1}{2} \quad \text{for} \quad i = 1, \dots, m \quad \text{and}$$

$$l = \lambda_1 + \dots + \lambda_m = \frac{m}{2\alpha}.$$

Using Part (2) of Lemma 9.1, we write

$$\mathbf{E} \exp\left\{ -\frac{m}{2\alpha} \sum_{j=1}^{k} v_j \right\} = \frac{1}{\#(R, C)} \sum_{D = (d_{ij})} \prod_{\substack{1 \leq i \leq m \\ 1 \leq j \leq k}} \frac{\Gamma(d_{ij} - \lambda_i + 1)}{\Gamma(d_{ij} + 1)},$$

where the sum is taken over all contingency tables D with margins (R, C).

The proof now follows by Proposition 11.1.   ∎

We will use the following corollary.

Corollary 12.2.   *Let (R, C) be lower β-smooth upper α-smooth margins such that $s = N/mn \geq 1$. Suppose further that $m \leq \rho n$ and $n \leq \rho m$ for some $\rho \geq 1$. Then for some $\tau = \tau(\alpha, \beta, \rho) > 0$ we have*

$$\mathbf{P}\{\#\{i : \ h_i \leq -\tau + \ln s\} > \ln N\} \leq 4^{-n} \quad and$$
$$\mathbf{P}\{\#\{j : \ v_j \leq -\tau + \ln s\} > \ln N\} \leq 4^{-m}.$$

Proof.   We introduce random sets

$$I = \{i : \ h_i \leq -\tau + \ln s\} \quad \text{and} \quad J = \{j : \ v_j \leq -\tau + \ln s\}$$

and note that

$$\frac{1}{\#I} \sum_{i \in I} h_i \leq -\tau + \ln s \quad \text{and} \quad \frac{1}{\#J} \sum_{j \in J} v_j \leq -\tau + \ln s.$$

The proof now follows from Lemma 12.1.   ∎

Proof of Theorem 3.2.   The proof is a modification of that of Theorem 3.1. We recall that

$$p(X) = \frac{N^N}{N!} \text{per } B(X),$$

where $B(X)$ is the $N \times N$ doubly stochastic matrix constructed as follows: we scale $m \times n$ matrix X to the matrix Y with row sums R and column sums C and let $b_{ij} = y_{pq}/r_p c_q$ provided the entry (i, j) lies in the (p, q)-th block $B(X)$ of size $r_p \times c_q$. We are going to bound the entries of Y. First, without loss of generality we assume that $s = N/mn \geq 1$ since the case of $s \leq 1$ is treated in Theorem 3.1.

As in the proof of Theorem 3.1 we conclude that

$$\mathbf{P}\left\{\frac{1}{N^2} \sum_{ij} r_i c_j x_{ij} < 2\alpha^2(s + 1)\right\} \geq 1 - \left(\frac{3}{4}\right)^{N+mn}. \tag{18}$$

Let

$$h_p = \frac{1}{N} \sum_{j=1}^{N} c_j \ln x_{pj} \quad \text{for} \quad p = 1, \ldots, m \quad \text{and}$$

$$v_q = \frac{1}{N} \sum_{i=1}^{m} r_i \ln x_{iq} \quad \text{for} \quad q = 1, \ldots, n.$$

Choose $\tau > 0$ as in Corollary 12.2. Set

$$P = \{p : \quad h_p \leq -\tau + \ln s\} \quad \text{and} \quad Q = \{q : \quad v_q \leq -\tau + \ln s\}.$$

Thus the probability that

$$\#P \le \ln N \quad \text{and} \quad \#Q \le \ln N$$

is at least

$$1 - 4^{-m} - 4^{-n}.$$

If $p \notin P$ and $q \notin Q$ and (18) holds then by Theorem 5.2,

$$y_{pq} \le \delta_1 \frac{r_p c_q}{sN} x_{pq}$$

for some $\delta_1(\alpha, \beta) > 0$. If $p \in P$ or $q \in Q$ then

$$y_{pq} \le \min\{r_p, c_q\}.$$

Consequently, for b_{ij} with (i,j) in the p, q-th block we have

$$b_{ij} \le \frac{\delta_1}{sN} x_{pq} \quad \text{if} \quad p \notin P \quad \text{and} \quad q \notin Q$$

and

$$b_{ij} \le \min\left\{\frac{1}{r_p}, \frac{1}{c_q}\right\} \quad \text{if} \quad p \in P \quad \text{or} \quad q \in Q.$$

As in the proof of Theorem 3.1, we let

$$z_i = \max_{j=1,\dots,N} b_{ij} \quad \text{for} \quad i = 1, \dots N \quad \text{and let}$$

$$u_p = \max_{q=1,\dots,m} x_{pq}.$$

We estimate that

$$z_i \le \frac{1}{r_p}$$

if i lies in the p-th row block with $p \in P$ and we estimate that

$$z_i \le \frac{\delta_1}{sN} u_p + \max_{q \in Q} \frac{y_{pq}}{r_p c_q},$$

if i lies in the row block $p \notin P$. Hence

$$\sum_{i=1}^{N} z_i \le \#P + \frac{\delta_1}{sN} \sum_{p=1}^{m} r_p u_p + \sum_{p=1}^{m} \max_{q \in Q} \frac{y_{pq}}{c_q}.$$

By Corollary 9.2,

$$\mathbf{P}\left\{\sum_{p=1}^{m} u_p \ge \tau_1 sm \ln N\right\} \le 4^{-m}$$

for some $\tau_1 = \tau_1(\alpha)$, and hence

$$\mathbf{P}\left\{\frac{\delta_1}{sN}\sum_{p=1}^{m} r_p u_p \leq \delta_2 \ln N\right\} \geq 1 - 4^{-m}.$$

for some $\delta_2 = \delta_2(\alpha)$. Finally,

$$\sum_{p=1}^{m} \max_{q \in Q} \frac{y_{pq}}{c_q} \leq \sum_{q \in Q}\sum_{p=1}^{m} \frac{y_{pq}}{c_q} \leq \delta_3 \#Q$$

for some $\delta_3 = \delta_3(\alpha, \beta)$. Summarizing,

$$\mathbf{P}\left\{\sum_{i=1}^{N} z_i \leq \delta \ln N\right\} \geq 1 - \left(\frac{3}{4}\right)^{N+mn} - 4^{-n} - 2 \cdot 4^{-m}$$

for some $\delta = \delta(\alpha, \beta, \rho) > 0$ and the proof is completed as in Theorem 3.1. ∎

ACKNOWLEDGMENTS

REFERENCES

[1] N. Alon and J. H. Spencer, The probabilistic method. With an appendix by Paul Erdös, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1992.

[2] D. Applegate and R. Kannan, "Sampling and integration of near log-concave functions," In Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing, ACM, New York, NY, 1991, pp. 156–163.

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

[4] R. B. Bapat and T. E. S. Raghavan, Nonnegative matrices and applications, Encyclopedia of mathematics and its applications 64, Cambridge University Press, Cambridge, 1997.

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

[6] A. Barvinok, Enumerating contingency tables via random permanents, Combinator Probabil Comput 17 (2008), 1–19.

[7] A. Barvinok, Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes, Int Math Res Notices 2009 (2009), 348–385.

[8] A. Barvinok, A. Samorodnitsky, and A. Yong, Counting magic squares in quasi-polynomial time, preprint arXiv:math/0703227, 2008.

[9] A. Békéssy, P. Békéssy, and J. Komlós, Asymptotic enumeration of regular matrices, Studia Scientiarum Math Hungarica 7 (1972), 343–353.

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

[11] L. M. Bregman, Certain properties of nonnegative matrices and their permanents (Russian), Doklady Akademii Nauk SSSR 211 (1973), 27–30, translated in Soviet Math Doklady 14 (1973), 945–949.

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

[13] Y. Chen, P. Diaconis, S. P. Holmes, and J. S. Liu, Sequential Monte Carlo methods for statistical analysis of tables, J Am Stat Assoc 100 (2005), 109–120.

[14] M. Cryan and M. Dyer, A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant, J Computer System Sci 67 (2003), 291–310.

[15] M. Cryan, M. Dyer, L. A. Goldberg, M. Jerrum, and R. Martin, Rapidly mixing Markov chains for sampling contingency tables with a constant number of rows, SIAM J Comput 36 (2006), 247–278.

[16] J. A. De Loera, R. Hemmecke, J. Tauzer, and R. Yoshida, Effective lattice point counting in rational convex polytopes, J Symbolic Computat 38 (2004), 1273–1302.

[17] 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 Stat 13 (1985), 845–913.

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

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

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

[21] D. I. Falikman, Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix (Russian), Matematicheskie Zametki 29 (1981), 931–938, translated in Math Notes 29 (1981), 475–479.

[22] S. Friedland, A lower bound for the permanent of a doubly stochastic matrix, Ann Math Second Series 110 (1979), 167–176.

[23] A. Frieze and R. Kannan, Log-Sobolev inequalities and sampling from log-concave distributions, Ann Appl Probabil 9 (1999), 14–26.

[24] A. Frieze, R. Kannan, and N. Polson, Sampling from log-concave distributions, Ann Appl Probabil 4 (1994), 812–837, correction, p. 1255.

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

[26] C. Greenhill and B. D. McKay, Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums, Adv Appl Math 41 (2008), 459–481.

[27] L. Gurvits, The van der Waerden conjecture for mixed discriminants, Adv Math 200 (2006), 435–454.

[28] L. Gurvits, Van der Waerden/Schrijver—Valiant like conjectures and stable (aka hyperbolic) homogenous polynomials: One theorem for all, Electronic J Combinator 15 (2008), Research Paper R66.

[29] M. Jerrum, A. Sinclair, and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, J ACM 51 (2004), 671–697.

[30] B. Kalantari and L. Khachiyan, On the complexity of nonnegative-matrix scaling, Linear Algebra Appl 240 (1996), 87–103.

[31] R. Kannan and S. Vempala, "Sampling lattice points," In STOC '97, El Paso, TX, ACM, New York, 1999, pp. 696–700.

[32] N. Linial, A. Samorodnitsky, and A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, Combinatorica 20 (2000), 545–568.

[33] L. Lovász and S. Vempala, "Fast algorithms for log-concave functions: Sampling, rounding, integration and optimization," In Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, IEEE Press, 2006, pp. 57–68.

[34] J. H. van Lint and R. M. Wilson, A course in combinatorics. Second edition, Cambridge University Press, Cambridge, 2001.

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

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

[37] 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.

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

[39] G. W. Soules, New permanental upper bounds for nonnegative matrices, Linear Multilinear Algebra 51 (2003), 319–337.

[40] S. Vempala, "Geometric random walks: A survey," In Combinatorial and computational geometry, Mathematical Sciences Research Institute Publications 52, Cambridge University Press, Cambridge, 2005, pp. 577–616.

[41] A. Yong, Contingency table and magic square enumeration, Software and data (2007), Available at: http://www.math.uiuc.edu/~ayong/papers.html. Accessed: December 17, 2009.