Combinatorics, Probability and Computing (2008) 17, 1–19. © 2007 Cambridge University Press
doi:10.1017/S0963548307008668 First published online 27 July 2007 Printed in the United Kingdom

Enumerating Contingency Tables via Random Permanents

ALEXANDER BARVINOK†

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043, USA
(e-mail: barvinok@umich.edu)

Received 29 November 2005; revised 4 March 2007

Given m positive integers $R = (r_i)$, n positive integers $C = (c_j)$ such that $\sum r_i = \sum c_j = N$, and mn non-negative weights $W = (w_{ij})$, we consider the total weight $T = T(R, C; W)$ of non-negative integer matrices $D = (d_{ij})$ with the row sums r_i, column sums c_j, and the weight of D equal to $\prod w_{ij}^{d_{ij}}$. For different choices of R, C, and W, the quantity $T(R, C; W)$ specializes to the permanent of a matrix, the number of contingency tables with prescribed margins, and the number of integer feasible flows in a network. We present a randomized algorithm whose complexity is polynomial in N and which computes a number $T' = T'(R, C; W)$ such that $T' \leqslant T \leqslant \alpha(R, C)T'$ where $\alpha(R, C) = \min\{\prod r_i! r_i^{-r_i}, \prod c_j! c_j^{-c_j}\} N^N / N!$. In many cases, $\ln T'$ provides an asymptotically accurate estimate of $\ln T$. The idea of the algorithm is to express T as the expectation of the permanent of an $N \times N$ random matrix with exponentially distributed entries and approximate the expectation by the integral T' of an efficiently computable log-concave function on $\mathbb{R}^{mn}$.

1. Introduction and main results

1.1. Contingency tables

Let us fix m positive integers $r_1, \ldots, r_m$ and n positive integers $c_1, \ldots, c_n$ such that

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

A non-negative $m \times n$ integer matrix $D = (d_{ij})$ with the row sums $r_1, \ldots, r_m$ and the column sums $c_1, \ldots, c_n$ is called a *contingency table* with the margins $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$. The problem of efficient enumeration of contingency tables with prescribed margins has attracted a lot of attention recently: see [7], [8], [10], [12] and [26]. The interest in contingency tables is motivated by applications to statistics, combinatorics, and representation theory: *cf.* [9] and [10].

† Research partially supported by NSF grant DMS 0400617. The author is grateful to Microsoft (Redmond) for hospitality during his work on this paper.

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Let $W = (w_{ij})$ be an $m \times n$ matrix of non-negative weights w_{ij}. In this paper, we consider the quantity (partition function)

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

where the sum is taken over all contingency tables $D = (d_{ij})$ with the given margins $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ and where we agree that $0^0 = 1$. Thus, if $w_{ij} = 1$ for all i and j, the value of $T(R, C; W)$ is equal to the number of contingency tables with the given margins. If $w_{ij} \in \{0, 1\}$, the number $T(R, C; W)$ counts contingency tables D for which we have $d_{ij} = 0$ for all i, j with $w_{ij} = 0$. In this case, $T(R, C; W)$ can be interpreted as the number of integer flows in a bipartite graph: see [3] and [8]. In Section 1.4 we show that counting integer feasible flows in an arbitrary network (with or without capacities on edges) reduces to computing $T(R, C; W)$ for some 0–1 matrix W. For certain matrices $W = (w_{ij})$ the problem of computing $T(R, C; W)$ naturally arises in statistics: *cf.* [11, Section 4.3].

Geometrically, one can view $T(R, C; W)$ as the generating function over all integer points in the transportation polytope of $m \times n$ non-negative matrices with the row sums r_i and column sums c_j: *cf.* [5].

We note that if $m = n$, $R = (1, \ldots, 1)$, and $C = (1, \ldots, 1)$ then

$$T(R, C; W) = \mathrm{per}(W)$$

is the *permanent* of the weight matrix W, that is,

$$\mathrm{per}(W) = \sum_\pi \prod_{i=1}^n w_{i\pi(i)},$$

where the sum is taken over all bijections $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$; *cf.*, for example, [25]. A randomized polynomial time approximation algorithm to compute the permanent of a given non-negative matrix was recently obtained by Jerrum, Sinclair and Vigoda [19].

We show that $T(R, C; W)$ can be represented as the expected permanent of an $N \times N$ random matrix with exponentially distributed entries.

Recall that a random variable γ is *standard exponential* if

$$\mathbf{P}\,(\gamma \geqslant \tau) = \begin{cases} e^{-\tau} & \text{for } \tau > 0, \\ 1 & \text{for } \tau \leqslant 0. \end{cases}$$

Our starting point is the following result.

Theorem 1.1. *Given positive integers $r_1, \ldots, r_m$ and $c_1, \ldots, c_n$ such that*

$$r_1 + \cdots + r_m = c_1 + \cdots + c_n = N,$$

and mn real numbers w_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, n$, let us construct the $N \times N$ random matrix A as follows. The matrix $A = A(\gamma)$ is a function of the $m \times n$ matrix $\gamma = (\gamma_{ij})$ of independent standard exponential random variables γ_{ij}. We represent the set of rows of A as a disjoint union of m subsets $R_1, \ldots, R_m$, where $|R_i| = r_i$ for $i = 1, \ldots, m$ and the set of columns of A as a disjoint union of n subsets $C_1, \ldots, C_n$, where $|C_j| = c_j$ for $j = 1, \ldots, n$.

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Thus A is split into mn blocks $R_i \times C_j$. We sample mn independent standard exponential random variables γ_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, n$, and fill the entries of the block $R_i \times C_j$ of $A = A(\gamma)$ by the copies $w_{ij}\gamma_{ij}$. Then the total weight $T(R, C; W)$ of the $m \times n$ contingency tables with the row sums $r_1, \ldots, r_m$ and column sums $c_1, \ldots, c_n$, where the table $D = (D_{ij})$ is counted with the weight

$$w(D) = \prod_{i,j} w_{ij}^{d_{ij}},$$

is equal to

$$\frac{\mathbf{E}\ \mathrm{per}(A)}{r_1! \cdots r_m! c_1! \cdots c_n!}.$$

We prove Theorem 1.1 in Section 2.

Although we can compute individual permanents via the algorithm of [19], evaluating the expectation is still a difficult problem. However, the expectation of an *approximate permanent* of A can be computed efficiently.

1.2. An approximation algorithm to compute $T(R, C; W)$

We rely heavily on the theory of *matrix scaling* and its applications to approximating the permanent, in particular as described in [17], [18], [20], [21], [23], [27], and [29], as well as on the Markov chain-based algorithms for integrating log-concave densities [2], [15], [16], [24] and [32].

We assume here that the weights w_{ij} are strictly positive, which is not really restrictive since the zero weights can be replaced by sufficiently small positive weights.

Let $A = (a_{ij})$ be an $N \times N$ positive matrix. Sinkhorn [29] proved that there exist positive numbers $\xi_1, \ldots, \xi_N; \eta_1, \ldots, \eta_N$ and a positive doubly stochastic (all row and column sums are equal to 1) $N \times N$ matrix $B = B(A)$, $B = (b_{ij})$, such that

$$a_{ij} = b_{ij}\xi_i\eta_j \quad \text{for} \quad i, j = 1, \ldots, N.$$

Moreover, the matrix $B = B(A)$ is unique, while the numbers $\xi_1, \ldots, \xi_N$ and $\eta_1, \ldots, \eta_N$ are unique up to a scaling $\xi_i \longmapsto \xi_i\tau$, $\eta_j \longmapsto \eta_j\tau^{-1}$ for some $\tau > 0$; see also Theorem 3.1. This allows us to define the function

$$\sigma(A) = \prod_{i=1}^{N} \xi_i\eta_i$$

(it is called the *capacity* of A in [17]). We use the following two crucial facts about σ.

- There is an algorithm which, given a positive $N \times N$ matrix A and a number $\epsilon > 0$, computes $\sigma(A)$ within relative error ϵ in time polynomial in $\ln \epsilon^{-1}$ and N [21].
- The function σ is *log-concave*, that is,

  $$\ln \sigma(\alpha_1 A_1 + \alpha_2 A_2) \geqslant \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2)$$

  for any positive matrices A_1 and A_2 and any non-negative α_1 and α_2 such that $\alpha_1 + \alpha_2 = 1$ (*cf.* [17]; see also Theorem 3.1).

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Our algorithm is based on replacing $\mathrm{per}(A)$ in Theorem 1.1 by the (scaled) capacity $\sigma(A)$. Namely, we define

$$T'(R,C;W) = \frac{N!}{N^N} \frac{\mathbf{E}\ \sigma(A)}{r_1!\cdots r_m! c_1!\cdots c_n!}.$$

Since both $\sigma(A)$ and the exponential density on $\mathbb{R}^{mn}$ are log-concave, and since $\sigma(A)$ is efficiently computable for any positive A, we can apply results of Kannan and colleagues [2, 15, 16] and of Lovász and Vempala [24, 32] on efficient integration of log-concave functions to show that there is a randomized fully polynomial time approximation scheme to compute $T'(R,C;W)$.

- We present a randomized algorithm which, for any $\epsilon > 0$, computes $T'(R,C;W)$ within relative error ϵ in time polynomial in ϵ^{-1} and N (in the unit cost model).

We discuss the details of our algorithm in Sections 3 and 4. Namely, in Section 3 we present the necessary results regarding $\sigma(A)$, while in Section 4 we discuss the integration problem.

Finally, we discuss how well the value of $T'(R,C;W)$ approximates $T(R,C;W)$.

Theorem 1.2. *For the number $T'(R,C;W)$ computed by the algorithm of Sections* 3 *and* 4*, we have*

$$T'(R,C;W) \leqslant T(R,C;W) \leqslant \alpha(R,C)T'(R,C;W),$$

where

$$\alpha(R,C) = \frac{N^N}{N!} \min\left\{ \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}},\quad \prod_{j=1}^{n} \frac{c_j!}{c_j^{c_j}} \right\}.$$

We prove Theorem 1.2 in Section 5.

1.3. Discussion

Using Stirling's formula

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

we conclude that

$$\alpha(R,C) = e^{O(m)}\sqrt{r_1\cdots r_m/N} \quad \text{and}$$
$$\alpha(R,C) = e^{O(n)}\sqrt{c_1\cdots c_n/N}.$$

Furthermore, for any fixed sum N, the largest value of $\alpha(R,C)$ is attained when all r_i and c_j are as close to each other as possible.

Let us fix m and n and let us scale $R \longrightarrow kR$ and $C \longrightarrow kC$ for some positive integer k. Suppose that $w_{ij} = 1$ for all i and j (we denote this matrix W by $\mathbf{1}$). Then the number $T(kR,kC;\mathbf{1})$ of contingency tables grows as a polynomial in k of degree $(m-1)(n-1)$; *cf.*, for example, [31, Chapter 4]. On the other hand, the approximation factor $\alpha(kR,kC)$ grows only as k^p for $p = \min\{(m-1)/2, (n-1)/2\}$. Thus, in this case, the algorithm of Section 1.2 allows us to capture the logarithmic order of $T(kR,kC;\mathbf{1})$.

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

It is quite plausible that $T'(R, C; \mathbf{1})$ approximates $T(R, C; \mathbf{1})$ much better than Theorem 1.2 suggests. After the first version of this paper was written, Alex Samorodnitsky [28] communicated to the author a proof that if $m = n$ and $R = C = (n, \ldots, n)$ then $T'(R, C; \mathbf{1})$ approximates $T(R, C; \mathbf{1})$ within a factor of $n^{O(\ln n)}$. Alexander Yong [33] implemented the algorithm of Sections 3 and 4 as a MAPLE code. Numerical experiments conducted by him [33] suggest that in the case of $R = C = (n, \ldots, n)$ the values of $T'(R, C; \mathbf{1})$ approximate $T(R, C; \mathbf{1})$ within a rather slowly growing, possibly even constant, factor.

Let us fix R, C, and W such that $T(R, C; W) > 1$. For a positive integer k, let us define a km-vector $R_k = (kR, \ldots, kR)$, consisting of k copies of the m-vector kR, a kn-vector $C_k = (kC, \ldots, kC)$, consisting of k copies of the n-vector kC, and a $km \times kn$ matrix W_k consisting of k^2 blocks W. It is not hard to show that

$$T(R_k, C_k; W_k) \geqslant T^{k^2}(R, C; W).$$

On the other hand, $\alpha(R_k, C_k)$ grows only as $k^{O(k)}$, so our algorithm allows us to capture the logarithmic order of $T(R_k, C_k; W_k)$ in this case as well. In particular, choosing $m = n$, $R = C = (l, \ldots, l)$, and $W = \mathbf{1}$, we can capture the logarithmic order of the number of $kn \times kn$ non-negative integer matrices ('magic squares'; see [31, Chapter 4]) with the row and column sums equal to kl, where n and l are fixed and k grows.

Apart from the case of $r_i = c_j = 1$ (computation of the permanent), where a fully polynomial randomized approximation scheme (FPRAS) was obtained by Jerrum, Sinclair and Vigoda [19], most of the research thus far dealt with the case of $w_{ij} = 1$, that is, with the non-weighted enumeration of contingency tables. Dyer, Kannan and Mount [12] showed that if the margins are not too small, $r_i = \Omega(n^2 m)$ and $c_j = \Omega(m^2 n)$, the Monte Carlo-based approach allows one to approximate the number of contingency tables within a prescribed relative error $\epsilon > 0$ in time polynomial in m, n, and ϵ^{-1}. In this case, the number of tables is well approximated by the volume of the transportation polytope of $m \times n$ non-negative matrices with the row sums r_i and the column sums c_j. Subsequently, Morris [26] obtained a similar result for the bounds $r_i = \Omega(n^{3/2} m \ln m)$ and $c_j = \Omega(m^{3/2} n \ln n)$. Cryan and Dyer [7] obtained a FPRAS for the number of tables when the number of rows (or columns) is fixed in advance. Subsequently, Cryan, Dyer and Randall [8] extended the result to counting contingency tables with bounds on the entries (again, either the number of rows or the number of columns is fixed in advance).

The approximation we get is much less precise, but applies to arbitrary weights $W = (w_{ij})$ and seems to be non-trivial even for $W = \mathbf{1}$ and moderate values of r_i, c_j. As we discussed above, in many non-trivial cases we get an asymptotically accurate estimate of $\ln T(R, C; W)$.

Since every log-concave density can be arbitrarily closely approximated by the push-forward (projection) of the Lebesgue measure restricted to some higher-dimensional convex body, the algorithm of Section 4 can be viewed as a volume approximation algorithm. In contrast to [12] and [26], the convex body whose volume we approximate is not polyhedral.

1.4. Counting integer flows in a graph

Let $G = (V, E)$ be a directed graph with the set V of vertices and the set E of edges. Hence every edge $e \in E$ is incident to the *head* $\mathrm{head}(v) \in V$ of e and the *tail* $\mathrm{tail}(e) \in V$.

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

We assume that G is connected and that it does not contain loops or multiple edges. Suppose further that each vertex v has an integer number $a(v)$, called the *excess* of v, assigned to it, and that

$$\sum_{v \in V} a(v) = 0.$$

A set of non-negative integers $x(e) : e \in E$ is called an *integer feasible flow* in G if, for every $v \in V$, the balance condition holds:

$$\sum_{e:\ \mathrm{head}(e)=v} x(e) - \sum_{e:\ \mathrm{tail}(e)=v} x(e) = a(v).$$

If G does not contain directed cycles $v_1 \to v_2 \to \cdots \to v_k \to v_1$, the number of integer feasible flows is finite, possibly 0. The problem of efficient counting of integer feasible flows in a given graph has attracted some attention recently: *cf.* [3] and [8]. A variation of the problem involves introducing capacities (upper bounds on the flows) on edges.

One can express the number of integer feasible flows in a graph with $|V| = n$ vertices as the number $T(R, C; W)$ of weighted $n \times n$ contingency tables, where $w_{ij} \in \{0, 1\}$ for all i, j. To this end, let us construct a bipartite graph with $n + n$ vertices as follows. For every vertex $v \in V$, we introduce the left copy v_L and the right copy v_R. The directed edges $u \to v$ of G are represented by the edges $u_L \to v_R$ of the bipartite graph. We also introduce edges $v_L \to v_R$. Finally, for every vertex $v \in V$ we choose an upper bound $z(v)$ for a possible inflow into v and assign the excesses by

$$a(v_L) = a(v) - z(v) \quad \text{and} \quad a(v_R) = z(v).$$

With an integer feasible flow in the original graph G we associate an integer feasible flow in the constructed bipartite graph. Namely, we let the flow on the edge $u_L \to v_R$ equal to the flow on the edge $u \to v$ for $u \neq v$ and we assign the flow on the edge $v_L \to v_R$ in a unique way to satisfy the constraints. This correspondence is a bijection between the integer feasible flows in G and the bipartite graph. Hence the number of such flows is equal to the number $T(R, C; W)$ of weighted $n \times n$ contingency tables with the rows and columns indexed by the vertices $v \in V$, the row margins $z(v) - a(v)$, the column margins $z(v)$ and the matrix $W = (w_{ij})$ of weights defined by $w_{ij} = 1$ if $(i, j) \in E$ or $i = j$ and $w_{ij} = 0$ for all other pairs (i, j).

Using a construction of [3] one can enforce inequalities $x(e) \leqslant c(e)$, where $c(e)$ are positive integer capacities on edges of G. For every capacitated edge $e = (u \to v)$ we introduce two new vertices u_1 and v_1 and replace the edge $u \to v$ by the three uncapacitated edges $u \to u_1$, $v_1 \to u_1$, and $v_1 \to v$ with the excesses $a(u_1) = c(e)$ and $a(v_1) = -c(e)$. As long as every directed cycle in G contains an edge with bounded capacity, the number of integer feasible flows in G remains finite and can be expressed as an appropriate functional $T(R, C; W)$.

2. Proof of Theorem 1.1

If γ is a standard exponential random variable, then for any integer $d \geqslant 0$ we have

$$\mathbf{E}\ \gamma^d = \int_0^{+\infty} \tau^d e^{-\tau}\ d\tau = d!.$$

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Let us consider the random matrix $A = (a_{pq})$ as defined in Theorem 1.1. We identify both the set of rows of A and the set of columns of A with the set $\{1, \ldots, N\}$.

For every permutation $\pi : \{1, \ldots, N\} \longrightarrow \{1, \ldots, N\}$, let

$$p_\pi = \prod_{k=1}^{N} a_{k\pi(k)} \tag{2.1}$$

be the corresponding term of per(A). Thus

$$\mathbf{E}\ \mathrm{per}(A) = \sum_{\pi} \mathbf{E}\ p_\pi, \tag{2.2}$$

where the sum is taken over all permutations π. With every permutation π we associate a contingency table $D = D(\pi)$, called the *pattern* of π, as follows. We let $D = (d_{ij})$, where d_{ij} is the number of indices $k \in \{1, \ldots, N\}$ such that $k \in R_i$ and $\pi(k) \in C_j$, so the $\big(k, \pi(k)\big)$th entry of A lies in the block $R_i \times C_j$ of A.

For the corresponding term p_π of the permanent (2.1), we have

$$\mathbf{E}\ p_\pi = \prod_{ij} w_{ij}^{d_{ij}} d_{ij}!, \tag{2.3}$$

where $D = (d_{ij})$ is the pattern of π.

Now, let us count how many permutations π have a given pattern $D = (d_{ij})$. Let us represent each subset R_i of rows as a disjoint (ordered) union

$$R_i = \bigcup_{j=1}^{n} R_{ij} \quad \text{for} \quad i = 1, \ldots, m$$

of (possibly empty) subsets R_{ij} with $|R_{ij}| = d_{ij}$, and each subset C_j of columns as a disjoint (ordered) union

$$C_j = \bigcup_{i=1}^{m} C_{ij} \quad \text{for} \quad j = 1, \ldots, n$$

of (possibly empty) subsets C_{ij} with $|C_{ij}| = d_{ij}$. This pair of partitions gives rise to exactly $\prod_{ij} d_{ij}!$ permutations π with the pattern D: we choose π in such a way that if $k \in R_{ij}$ then $\pi(k) \in C_{ij}$, and we note that there are precisely $d_{ij}!$ bijections $R_{ij} \longrightarrow C_{ij}$.

On the other hand, the number of partitions $R_i = \bigcup_j R_{ij}$ is

$$\frac{r_i!}{\prod_{j=1}^{n} d_{ij}!}$$

while the number of partitions $C_j = \bigcup_i C_{ij}$ is

$$\frac{c_j!}{\prod_{i=1}^{m} d_{ij}!}.$$

Therefore, the number of permutations with the given pattern $D = (d_{ij})$ is

$$\frac{r_1! \cdots r_m! c_1! \cdots c_n!}{\prod_{ij} d_{ij}!}.$$

The proof now follows by (2.3) and (2.2). □

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Remark. Let us modify the definition of A as follows: instead of filling the $R_i \times C_j$ block with copies of $w_{ij}\gamma_{ij}$, we fill $R_i \times C_j$ with copies of just w_{ij}, so A is constructed deterministically. It follows from the proof above that the value of

$$\frac{\operatorname{per}(A)}{r_1! \cdots r_m! c_1! \cdots c_n!}$$

is equal to the total weight of contingency tables with the margins $r_1, \ldots, r_m$ and $c_1, \ldots, c_n$, provided the weight of the table $D = (d_{ij})$ is

$$\prod_{ij} \frac{w_{ij}^{d_{ij}}}{d_{ij}!}$$

(the Fisher–Yates statistics; *cf.* [10]).

For another proof of Theorem 1.1 in a particular case of $w_{ij} = 1$, see [4].

3. Matrix scaling

Here we summarize the matrix scaling results that we need. All the results in Theorem 3.1 below can be found in the literature, although sometimes stated in different terms.

We reproduce the approach of Gurvits and Samorodnitsky [18] adapted to the case of the scaling of a single positive matrix (paper [18] treats a more general and more complicated scaling of an N-tuple of $N \times N$ positive definite matrices). The approach of [18] is, in turn, a modification of London's approach [23].

Theorem 3.1. *For every positive $N \times N$ matrix $A = (a_{ij})$, there exist unique positive N-vectors $x = x(A)$, $y = y(A)$, and an $N \times N$ positive matrix $B = B(A)$*

$$x = (\xi_1, \ldots, \xi_N), \quad y = (\eta_1, \ldots, \eta_N), \quad \text{and} \quad B = (b_{ij})$$

so that the following holds.

(1) *We have*

$$a_{ij} = b_{ij}\xi_i\eta_j \quad \text{for} \quad i, j = 1, \ldots, N.$$

(2) *We have*

$$\prod_{j=1}^{N} \eta_j = 1.$$

(3) *Matrix B is doubly stochastic, that is,*

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

Let us define

$$\sigma(A) = \prod_{i=1}^{N} \xi_i, \quad \text{where} \quad x(A) = (\xi_1, \ldots, \xi_N).$$

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Then function σ satisfies the following properties on the set of positive $N \times N$ matrices.

(4) ***Positive homogeneity.*** *We have*

$$\sigma(\lambda A) = \lambda^N \sigma(A)$$

for any positive $N \times N$ matrix A and $\lambda > 0$.

(5) ***Monotonicity.*** *If $A = (a_{ij})$ and $A' = (a'_{ij})$ are positive $N \times N$ matrices such that $a'_{ij} \leqslant a_{ij}$ for all i, j, then*

$$\sigma(A') \leqslant \sigma(A).$$

(6) ***Log-concavity.*** *We have*

$$\ln \sigma(\alpha_1 A_1 + \alpha_2 A_2) \geqslant \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2)$$

for any two positive $N \times N$ matrices A_1 and A_2 and any two numbers $\alpha_1, \alpha_2 \geqslant 0$ such that $\alpha_1 + \alpha_2 = 1$.

Proof. Let us consider the hyperplane

$$H = \left\{ (\tau_1, \ldots, \tau_N) : \quad \sum_{i=1}^{N} \tau_i = 0 \right\}$$

in $\mathbb{R}^N$. With a positive matrix $A = (a_{ij})$, we associate a function $f_A : \mathbb{R}^N \longrightarrow \mathbb{R}$,

$$f_A(t) = \sum_{i=1}^{N} \ln \left(\sum_{j=1}^{N} a_{ij} e^{\tau_j} \right), \quad \text{where} \quad t = (\tau_1, \ldots, \tau_N).$$

Then the restriction of $f_A(t)$ on H is strictly convex and, moreover, f_A attains its unique minimum $t^* = (\tau_1^*, \ldots, \tau_N^*)$, $t^* = t^*(A)$, on H: see [18]. To relate our approach to [18], we note that a positive $N \times N$ matrix A gives rise to an N-tuple of $N \times N$ positive definite matrices D_i, where each D_i is the diagonal matrix having the ith row of A as its diagonal.

Since f_A is smooth, t^* is also a critical point and the gradient of f_A at the critical point is proportional to vector $(1, \ldots, 1)$, from which we get

$$\sum_{i=1}^{N} \left(\frac{a_{ik} e^{\tau_k^*}}{\sum_{j=1}^{N} a_{ij} e^{\tau_j^*}} \right) = \gamma \tag{3.1}$$

for some constant γ and $k = 1, \ldots, N$.

Let

$$\xi_i = \sum_{j=1}^{N} a_{ij} e^{\tau_j^*} \quad \text{for} \quad i = 1, \ldots, N,$$

let

$$\eta_j = e^{-\tau_j^*} \quad \text{for} \quad j = 1, \ldots, N,$$

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

and let us define an $N \times N$ matrix $B = (b_{ij})$ by

$$b_{ij} = \frac{a_{ij}}{\xi_i \eta_j} \quad \text{for} \quad i, j = 1, \ldots, N.$$

We note that

$$\prod_{j=1}^{N} \eta_j = 1$$

since t^* lies in the hyperplane H with $\tau_1 + \cdots + \tau_N = 0$. Then, by (3.1), we have

$$\sum_{i=1}^{N} b_{ij} = \gamma \quad \text{for} \quad j = 1, \ldots, N. \tag{3.2}$$

On the other hand,

$$\sum_{j=1}^{N} b_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, N. \tag{3.3}$$

Since B is a square matrix, comparing (3.2) and (3.3), we infer that $\gamma = 1$ and therefore we established the existence of $x = (\xi_1, \ldots, \xi_N)$ and $y = (\eta_1, \ldots, \eta_N)$ and B satisfying parts (1)–(3) of the theorem.

To show uniqueness, we note that if $x = (\xi_1, \ldots, \xi_N)$, $y = (\eta_1, \ldots, \eta_N)$, and B satisfy parts (1)–(3), then we must have

$$\xi_i = \sum_{j=1}^{N} a_{ij} \eta_j^{-1} \quad \text{for} \quad i = 1, \ldots, N$$

and hence, necessarily, the point $t = (\tau_1, \ldots, \tau_N)$ defined by

$$\tau_j = -\ln \eta_j \quad \text{for} \quad j = 1, \ldots, N$$

is a critical point of $f_A(t)$ on H. Since $f_A(t)$ is strictly convex on H, there is a unique critical point $t^* = t^*(A)$.

Thus function $\sigma(A)$ is well-defined. Part (4) follows since, by uniqueness of the scaling factors, we must have

$$x(\lambda A) = \lambda x(A) \quad \text{and} \quad y(\lambda A) = y(A).$$

To prove parts (5) and (6), we note that, by our construction,

$$\ln \sigma(A) = \sum_{i=1}^{N} \ln \xi_i = f_A(t^*) = \min_{t \in H} f_A(t). \tag{3.4}$$

Since $f_{A'}(t) \leqslant f_A(t)$ for all t, part (5) follows by (3.4). Finally, we observe that for any fixed t, the function $g(A) = f_A(t)$ is concave on the set of positive matrices $A = (a_{ij})$. Hence for any $t \in H$ and any $\alpha_1, \alpha_2 \geqslant 0$ such that $\alpha_1 + \alpha_2 = 1$, we have

$$f_{\alpha_1 A_1 + \alpha_2 A_2}(t) \geqslant \alpha_1 f_{A_1}(t) + \alpha_2 f_{A_2}(t) \geqslant \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2).$$

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Taking the minimum over $t \in H$, we conclude by (3.4) that

$$\ln \sigma(\alpha_1 A_1 + \alpha_2 A_2) \geqslant \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2),$$

and part (6) follows. □

3.1. Computing $\sigma(A)$

Linial, Samorodnitsky and Wigderson present in [21] a deterministic polynomial time algorithm, which, given an $N \times N$ positive matrix A and a number $\epsilon > 0$, computes the value of $\sigma(A)$ within a factor of $(1+\epsilon)$ in time polynomial in $\ln \epsilon^{-1}$ and N (in the unit cost model). Moreover, their algorithm extends to the scaling of a positive (not necessarily square) matrix to the given row and column sums.

We are interested in computing $\sigma(A)$ where $A = A(\gamma)$ is a random matrix of Theorem 1.1. Thus A is positive with probability 1. The special block structure of A allows us to save on computations as follows.

Let us consider the $m \times n$ matrix $(w_{ij}\gamma_{ij})$, which is also positive with probability 1. Applying the algorithm of [21], we scale the matrix to the row sums r_i and the column sums c_j. Namely, we compute (approximately, in polynomial time) positive numbers λ_i, $i = 1, \ldots, m$, and μ_j, $j = 1, \ldots, n$, and an $m \times n$ positive matrix $L = (l_{ij})$ such that

$$w_{ij}\gamma_{ij} = l_{ij}\mu_i\lambda_j \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n$$

and such that

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

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

If we divide every row of A from R_i by $\mu_i r_i$ and divide every column from C_j by $\lambda_j c_j$, we get the $N \times N$ matrix with the entries in the $R_i \times C_j$ block equal to $l_{ij}/r_i c_j$. It is seen that the obtained matrix is doubly stochastic. Therefore, we have

$$\sigma(A) = \left(\prod_{i=1}^{m} (\mu_i r_i)^{r_i} \right) \left(\prod_{j=1}^{n} (\lambda_j c_j)^{c_j} \right).$$

Hence the scaling of the $N \times N$ matrix A reduces to the scaling of the (generally speaking, much smaller) $m \times n$ matrix $(w_{ij}\gamma_{ij})$. In the implementation of the algorithm [33], the classical Sinkhorn balancing (repeated alternate scaling of rows and columns to the desired sums; see [29]) was employed and turned out to work just fine for $m, n \leqslant 100$.

4. Integrating $\sigma(A)$

Here we describe an algorithm for computing

$$T'(R, C; W) = \frac{N!}{N^N} \frac{\mathbf{E}\ \sigma(A)}{r_1! \cdots r_m! c_1! \cdots c_n!}$$

(*cf.* Section 3).

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

4.1. Notation

We interpret the space $\mathbb{R}^{mn}$ as the space of all $m \times n$ matrices $\gamma = (\gamma_{ij})$. Let $\mathbb{R}^{mn}_+$ denote the positive orthant $\gamma_{ij} > 0$ of $\mathbb{R}^{mn}$, and let

$$\Delta = \left\{\gamma : \quad \sum_{ij} \gamma_{ij} = 1 \quad \text{and} \quad \gamma_{ij} > 0 \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n\right\}$$

be the standard (open) simplex in $\mathbb{R}^{mn}$.

For $0 < \delta < 1/mn$ let us consider the *δ-interior* of Δ:

$$\Delta_\delta = \left\{\gamma : \quad \sum_{ij} \gamma_{ij} = 1 \quad \text{and} \quad \gamma_{ij} > \delta \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n\right\}.$$

Geometrically, Δ_δ is an open simplex lying strictly inside Δ.

For $\tau \in \mathbb{R}$, let us consider the set of parallel affine hyperplanes $H_\tau \subset \mathbb{R}^{mn}$:

$$\sum_{ij} \gamma_{ij} = \tau.$$

Let ν be the Lebesgue measure on the hyperplane H_1 normalized by the condition $\nu(\Delta) = 1$. Abusing notation, we also denote by ν the measure on H_τ that is obtained by a translation of the measure on H_1.

For a matrix $\gamma \in \mathbb{R}^{mn}_+$, let

$$P(\gamma) = \operatorname{per}\big(A(\gamma)\big), \quad \text{and let} \quad S(\gamma) = \sigma\big(A(\gamma)\big),$$

where $A(\gamma)$ is the matrix constructed in Theorem 1.1 and σ is the function of Theorem 3.1.

Thus we have

$$T(R, C; W) = \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \int_{\mathbb{R}^{mn}_+} P(\gamma) \exp\left\{-\sum_{ij} \gamma_{ij}\right\} d\gamma$$

and

$$T'(R, C; W) = \frac{N!}{N^N} \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \int_{\mathbb{R}^{mn}_+} S(\gamma) \exp\left\{-\sum_{ij} \gamma_{ij}\right\} d\gamma,$$

where $d\gamma$ is the Lebesgue measure on $\mathbb{R}^{mn}$.

To apply the results of [2], [15], [16], [24] and [32] on efficient integration of log-concave functions, we modify the problem to that of integration of $P(\gamma)$ and $S(\gamma)$, first on Δ and then on Δ_δ.

We use that both functions $P(\gamma)$ and $S(\gamma)$ are positive homogeneous of degree N and monotone on $\mathbb{R}^{mn}_+$: if $\gamma = (\gamma_{ij})$ and $\gamma' = (\gamma'_{ij})$ are positive matrices such that

$$\gamma'_{ij} \leqslant \gamma_{ij} \quad \text{for all} \quad i, j,$$

then

$$P(\gamma') \leqslant P(\gamma) \quad \text{and} \quad S(\gamma') \leqslant S(\gamma)$$

(*cf.* Theorem 3.1).

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Lemma 4.1. *We have*

$$\int_{\mathbb{R}_+^{mn}} P(\gamma) \exp\left\{-\sum_{ij} \gamma_{ij}\right\} d\gamma = \frac{(N+mn-1)!}{(mn-1)!} \int_\Delta P(\gamma)\ dv(\gamma)$$

and

$$\int_{\mathbb{R}_+^{mn}} S(\gamma) \exp\left\{-\sum_{ij} \gamma_{ij}\right\} d\gamma = \frac{(N+mn-1)!}{(mn-1)!} \int_\Delta S(\gamma)\ dv(\gamma).$$

Proof. We note that

$$\mathbb{R}_+^{mn} = \bigcup_{\tau>0} \tau\Delta.$$

Since $dv\ d\tau = (mn-1)!\ d\gamma$, we get

$$\int_{\mathbb{R}_+^{mn}} P(\gamma) \exp\left\{-\sum_{ij} \gamma_{ij}\right\} d\gamma = \frac{1}{(mn-1)!} \int_0^{+\infty} e^{-\tau} \left(\int_{\tau\Delta} P(\gamma)\ dv(\gamma) \right) d\tau.$$

Since $P(\gamma)$ is positive homogeneous of degree N, we conclude that

$$\int_{\tau\Delta} P(\gamma)\ dv(\gamma) = \tau^{N+mn-1} \int_\Delta P(\gamma)\ dv(\gamma),$$

which proves the first identity. The second identity is proved similarly. □

Next, we approximate the integral over the simplex Δ by the integral over the inner simplex Δ_δ.

Lemma 4.2. *Let* $0 < \delta < 1/mn$. *Then*

$$\big(1-mn\delta\big)^{N+mn-1} \int_\Delta P(\gamma)\ dv(\gamma) \leqslant \int_{\Delta_\delta} P(\gamma)\ dv(\gamma) \leqslant \int_\Delta P(\gamma)\ dv(\gamma)$$

and

$$\big(1-mn\delta\big)^{N+mn-1} \int_\Delta S(\gamma)\ dv(\gamma) \leqslant \int_{\Delta_\delta} S(\gamma)\ dv(\gamma) \leqslant \int_\Delta S(\gamma)\ dv(\gamma).$$

Proof. We observe that the translation

$$\gamma_{ij} \longmapsto \gamma_{ij} - \delta \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n$$

maps Δ_δ inside $(1-\delta mn)\Delta$. Since P is monotone and positive homogeneous of degree N, we get

$$\int_{\Delta_\delta} P(\gamma)\ dv(\gamma) \geqslant \int_{(1-\delta mn)\Delta} P(\gamma)\ dv(\gamma) = (1-\delta mn)^{N+mn-1} \int_\Delta P(\gamma)\ dv(\gamma).$$

The corresponding inequality for the integrals of $S(\gamma)$ is proved in a similar way and the remaining inequalities are obvious. □

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

For an $0 < \epsilon < 1$, let us choose a positive

$$\delta \leqslant \frac{-\ln(1-\epsilon)}{mn(N+mn-1)} \approx \frac{\epsilon}{mn(N+mn-1)} \quad \text{for small} \quad \epsilon > 0.$$

Then the integrals

$$\int_{\Delta_\delta} P(\gamma)\, dv(\gamma) \quad \text{and} \quad \int_{\Delta_\delta} S(\gamma)\, dv(\gamma)$$

approximate the respective integrals

$$\int_{\Delta} P(\gamma)\, dv(\gamma) \quad \text{and} \quad \int_{\Delta} S(\gamma)\, dv(\gamma)$$

within a factor of $(1-\epsilon)$.

Since $A(\gamma)$ depends linearly on γ, by the results of Section 3, $S(\gamma)$ is a strictly positive log-concave function on the set of positive matrices γ and the value of $S(\gamma)$ can be computed in polynomial time for any given positive matrix γ.

Our goal consists of estimating $T(R, C; W)$ by

$$T'_\delta(R, C; W) = \frac{N!}{N^N} \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \frac{(N+mn-1)!}{(mn-1)!} \int_{\Delta_\delta} S(\gamma)\, dv(\gamma).$$

To compute the integral, we apply one of the algorithms of [2], [16], [15] and [24] (see also [32] for a survey). The computational complexity of the algorithms is polynomial in the dimension $mn-1$ of the integral *and* the Lipschitz constant of $\ln S$ on Δ_δ. Hence it remains to estimate the Lipschitz constant of $\ln S$.

Lemma 4.3. *Let $0 < \delta < 1/mn$ be a number. Let $\gamma = (\gamma_{ij})$ and $\gamma' = (\gamma'_{ij})$ be two matrices such that*

$$\gamma_{ij}, \gamma'_{ij} \geqslant \delta \quad \textit{for all} \quad i, j.$$

Then

$$|\ln S(\gamma) - \ln S(\gamma')| \leqslant \frac{N}{\delta} \max_{ij} |\gamma_{ij} - \gamma'_{ij}|.$$

Proof. Let

$$\alpha = \max_{ij} |\gamma_{ij} - \gamma'_{ij}|.$$

Then

$$\gamma_{ij} \leqslant \gamma'_{ij} + \alpha \leqslant \gamma'_{ij}\left(1 + \frac{\alpha}{\delta}\right) \quad \text{for all} \quad i, j$$

and, similarly,

$$\gamma'_{ij} \leqslant \gamma_{ij} + \alpha \leqslant \gamma_{ij}\left(1 + \frac{\alpha}{\delta}\right) \quad \text{for all} \quad i, j.$$

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Since function $S(\gamma)$ is monotone and positive homogenous of degree N, we have

$$\ln S(\gamma) \leqslant \ln S(\gamma') + N \ln\left(1 + \frac{\alpha}{\delta}\right) \leqslant \ln S(\gamma') + \frac{\alpha N}{\delta}$$

and, similarly,

$$\ln S(\gamma') \leqslant \ln S(\gamma) + N \ln\left(1 + \frac{\alpha}{\delta}\right) \leqslant \ln S(\gamma) + \frac{\alpha N}{\delta},$$

which completes the proof. □

Summarizing, we conclude that there is a randomized algorithm which, for any given $\epsilon > 0$ computes the value of

$$T'(R, C; W) = \frac{N!}{N^N} \frac{\mathbf{E}\ \sigma(A)}{r_1! \cdots r_m! c_1! \cdots c_n!},$$

where A is the matrix of Theorem 1.1, within relative error ϵ in time polynomial in ϵ^{-1} and N (in the unit cost model).

The implementation of [33] uses a version of the 'hit-and-run' algorithm of [24]; see also [32].

5. Proof of Theorem 1.2

Our proof is based on two estimates for the permanent of a non-negative matrix.

5.1. The van der Waerden bound

Let $B = (b_{ij})$ be an $N \times N$ doubly stochastic matrix, that is, a non-negative matrix with all row and column sums equal to 1. Then

$$\operatorname{per}(B) \geqslant \frac{N!}{N^N}.$$

This bound constituted the van der Waerden conjecture proved by Egorychev [13] and Falikman [14]; see also [22, Chapter 12].

5.2. A continuous extension of the Minc–Bregman bound

Let $B = (b_{ij})$ be an $N \times N$ non-negative matrix. Let

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

be the row sums of B. If $b_{ij} \in \{0, 1\}$, the bound

$$\operatorname{per}(B) \leqslant \prod_{i=1}^{N} (s_i!)^{1/s_i}$$

was conjectured by Minc and proved by Bregman [6]; see also [1, Chapter 2].

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

Alex Samorodnitsky [28] communicated to the author the following extension of the Minc–Bregman bound. Suppose that

$$\sum_{j=1}^{N} b_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, N \tag{5.1}$$

and that

$$b_{ij} \leqslant 1/t_i \quad \text{for all} \quad i = 1, \ldots, N, \tag{5.2}$$

where $t_i, i = 1, \ldots, N$, are positive integers. Then

$$\operatorname{per}(B) \leqslant \prod_{i=1}^{N} \frac{(t_i!)^{1/t_i}}{t_i}. \tag{5.3}$$

To deduce (5.3), we argue that the maximum of $\operatorname{per}(B)$ on the class of $N \times N$ non-negative matrices satisfying (5.1) and (5.2) is attained at a matrix with $b_{ij} \in \{0, 1/t_i\}$ for all i, j. Indeed, let us choose a particular row index i. Then any non-negative matrix B satisfying (5.1)–(5.2) can be written as a convex combination of two non-negative matrices B' and B'' which satisfy (5.1)–(5.2), agree with B in all rows, except possibly the ith row, and, additionally, satisfy $b'_{ij}, b''_{ij} \in \{0, 1/t_i\}$. Since the function $\operatorname{per}(B)$ is linear in every row, we conclude that $\operatorname{per}(B) \leqslant \max\{\operatorname{per}(B'), \operatorname{per}(B'')\}$. Proceeding as above for rows $i = 1, \ldots, N$, we may assume that $b_{ij} \in \{0, 1/t_i\}$ for all i, j, so (5.3) follows from the Minc–Bregman bound.

A similar bound is obtained by Soules [30]. If B is a non-negative matrix satisfying (5.1)–(5.2) where t_i need not be integers, then

$$\operatorname{per}(B) \leqslant \prod_{i=1}^{N} \frac{\Gamma^{1/t_i}(t_i + 1)}{t_i}.$$

Now we are ready to complete the proof of Theorem 1.2.

Proof. For a given $m \times n$ positive matrix $\gamma = (\gamma_{ij})$, let $A = A(\gamma)$ be the matrix constructed in Theorem 1.1 and let $B = B(A)$ be the matrix constructed in Theorem 3.1. We have

$$\operatorname{per}(A) = \sigma(A)\operatorname{per}(B),$$

and we estimate $\operatorname{per}(B)$.

By the estimate of Section 5.1, we have

$$\operatorname{per}(A) \geqslant \frac{N!}{N^N}\sigma(A),$$

from which we obtain

$$T(R, C; W) \geqslant T'(R, C; W).$$

On the other hand, as is discussed in Section 3.1, we can construct $B = B(\gamma)$ as follows. First, we construct a positive $m \times n$ matrix $L = (l_{ij})$ such that

$$w_{ij}\gamma_{ij} = l_{ij}\mu_i\lambda_j \quad \text{for all} \quad i, j$$

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

for some positive numbers $\mu_1, \ldots, \mu_m$ and $\lambda_1, \ldots, \lambda_n$, and such that

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

and

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

Then we fill the $R_i \times C_j$ block of B by $l_{ij}/r_i c_j$.

It follows then that every entry of B in the block R_i of rows does not exceed $1/r_i$. Applying the bound of Section 5.2, we get

$$\operatorname{per}(B) \leqslant \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}.$$

Similarly, every entry of B in the block C_j of columns does not exceed $1/c_j$, so we get

$$\operatorname{per}(B) \leqslant \prod_{i=1}^{n} \frac{c_j!}{c_j^{c_j}}.$$

Therefore,

$$\operatorname{per}(A) \leqslant \sigma(A) \min \left\{ \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}, \quad \prod_{i=1}^{n} \frac{c_j!}{c_j^{c_j}} \right\}.$$

and the proof follows. □

Conclusion

We defined the quantity (partition function) $T(R, C; W)$, which for various values of the parameters specializes to the permanent of a non-negative matrix, the number of contingency tables with prescribed margins, and the number of integer feasible flows in a network. It has some applications to statistics and is interesting in its own right. We presented a randomized polynomial time algorithm which approximates $T(R, C; W)$ within a certain factor $\alpha(R, C)$. Although $\alpha(R, C)$ is exponentially large, in many interesting cases the approximation gives us the correct order of $\ln T(R, C; W)$. It is quite plausible that in some particular situations, such as the contingency table counting, the approximation that we actually obtain is much better than that we are able to prove. The algorithm approximates $T(R, C; W)$ by the integral of an efficiently computable log-concave function and the integral is computed by known randomized polynomial time algorithms. The algorithm has been implemented.

Acknowledgements

The author is grateful to Leonid Gurvits for useful conversations and references, to Alex Samorodnitsky, whose communication [28] resulted in sharper bounds in Theorem 1.2, and to Alexander Yong for implementing the algorithm.

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

References

[1] Alon, N. and Spencer, J. (2000) *The Probabilistic Method*, 2nd edn, Wiley–Interscience.

[2] Applegate, D. and Kannan, R. (1991) Sampling and integration of log-concave functions. In *Proc. Twenty-Third Annual ACM Symposium on Theory of Computing*, ACM Press, pp. 156–163.

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

[4] Barvinok, A. (2005) Low rank approximations of symmetric polynomials and asymptotic counting of contingency tables. Preprint: `arXiv math.CO/0503170`.

[5] Barvinok, A. and Pommersheim, J. E. (1999) An algorithmic theory of lattice points in polyhedra. In *New Perspectives in Algebraic Combinatorics* (Berkeley, CA, 1996–97), Vol. 38 of *Math. Sci. Res. Inst. Publ.*, Cambridge University Press, pp. 91–147.

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

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

[8] Cryan, M., Dyer, M. and Randall, D. (2005) Approximately counting integral flows and cell-bounded contingency tables. In *STOC'05: Proc. 37th Annual ACM Symposium on Theory of Computing*, ACM Press, pp. 413–422.

[9] Diaconis, P. and Gamburd, A. (2004) Random matrices, magic squares and matching polynomials. *Electron. J. Combin.* **11** # 2.

[10] Diaconis, P. and Gangolli, A. (1995) Rectangular arrays with fixed margins. In *Discrete Probability and Algorithms* (Minneapolis, MN, 1993), Vol. 72 of *IMA Vol. Math. Appl.*, Springer, pp. 15–41.

[11] Diaconis, P. and Sturmfels, B. (1998) Algebraic algorithms for sampling from conditional distributions. *Annals Statist.* **26** 363–397.

[12] Dyer, M., Kannan, R. and Mount, J. (1997) Sampling contingency tables. *Random Struct. Alg.* **10** 487–506.

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

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

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

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

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

[18] Gurvits, L. and Samorodnitsky, A. (2002) A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary. *Discrete Comput. Geom.* **27** 531–550.

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

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

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

[22] van Lint, J. H. and Wilson, R. M. (2001) *A Course in Combinatorics*, 2nd edn, Cambridge University Press.

[23] London, D. (1971) On matrices with a doubly stochastic pattern. *J. Math. Anal. Appl.* **34** 648–652.

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press

[24] Lovász, L. and Vempala, S. (2006) Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *Proc. 47th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Press, pp. 57–68.
[25] Minc, H. (1978) *Permanents*, Addison-Wesley.
[26] Morris, B. J. (2002) Improved bounds for sampling contingency tables. *Random Struct. Alg.* **21** 135–146.
[27] Nemirovski, A. and Rothblum, U. (1999) On complexity of matrix scaling. *Linear Algebra Appl.* **302/303** 435–460.
[28] Samorodnitsky, A. (2006) Personal communication.
[29] Sinkhorn, R. (1964) A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.* **35** 876–879.
[30] Soules, G. W. (2003) New permanental upper bounds for nonnegative matrices. *Linear Multilinear Algebra* **51** 319–337.
[31] Stanley, R. P. (1997) *Enumerative Combinatorics*, Vol. 1. Corrected reprint of the 1986 original, Cambridge University Press.
[32] Vempala, S. (2005) Geometric random walks: A survey. In *Combinatorial and Computational Geometry*, Vol. 52 of *Mathematical Sciences Research Institute Publications*, Cambridge University Press, pp. 577–616.
[33] Yong, A. (2006) Contingency table and magic square enumeration. Available at: `http://www.math.umn.edu/~ayong/contingency.html`

https://doi.org/10.1017/S0963548307008668 Published online by Cambridge University Press