

On the mixing time of the Diaconis–Gangolli random walk on contingency tables over $\mathbb{Z}/q\mathbb{Z}$

Evita Nestoridi and Oanh Nguyen

Department of Mathematics, Princeton University, USA. E-mail: exn@princeton.edu; onguyen@princeton.edu

Received 15 October 2018; revised 12 February 2019; accepted 2 April 2019

Abstract. The Diaconis–Gangolli random walk is an algorithm that generates an almost uniform random graph with prescribed degrees. In this paper, we study the mixing time of the Diaconis–Gangolli random walk restricted on $n \times n$ contingency tables over $\mathbb{Z}/q\mathbb{Z}$. We prove that the random walk exhibits cutoff at $\frac{n^2}{4(1-\cos(\frac{2\pi}{q}))} \log n$, when $\log q = o(\frac{\sqrt{\log n}}{\log \log n})$.

Résumé. La marche aléatoire de Diaconis–Gangolli est un algorithme qui génère un graphe aléatoire à degrés prescrits, de loi presque uniforme. Dans cet article, nous étudions le temps de mélange de cette marche aléatoire restreinte aux tableaux de contingence de taille $n \times n$ sur $\mathbb{Z}/q\mathbb{Z}$. Nous montrons que la marche aléatoire présente une transition abrupte (cutoff) à $\frac{n^2}{4(1-\cos(\frac{2\pi}{q}))} \log n$, où $\log q = o(\frac{\sqrt{\log n}}{\log \log n})$.

MSC: Primary 60J10; secondary 60C05; 60G42

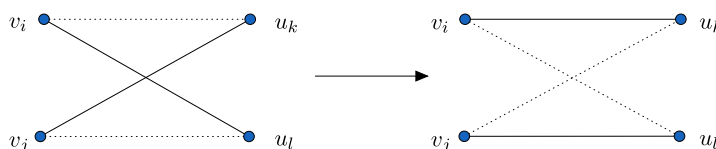
Keywords: Contingency table; Random graphs; Mixing time; Cutoff

1. Introduction

1.1. Motivation and main result

Random graphs are an important object of study in Combinatorics, Computer Science, and Probability. An (n, n) bipartite graph is a graph with a left vertex set $V = \{v_1, \dots, v_n\}$ and a right vertex set $U = \{u_1, \dots, u_n\}$ and the only edges are the edges that connect a vertex in V with a vertex in U . Consider the problem of generating a random bipartite graph with prescribed degrees. How do we choose uniformly at random such an (n, n) bipartite graph?

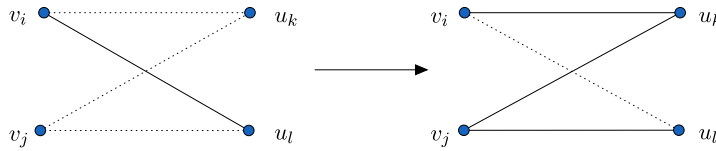
Diaconis and Gangolli [30] proposed the following randomized algorithm. Start with a bipartite graph G_0 that has the desired vertex degrees at time $t = 0$. At time t , from the graph G_{t-1} , choose two pairs of vertices $v_i \neq v_j$ in V and $u_k \neq u_l$ in U uniformly at random. We delete the already existing edges between v_i, v_j and u_k, u_l and we draw the remaining edges between them to get G'_t . If the result is not a graph with the prescribed degrees, then G_t is set to be G_{t-1} . Otherwise, $G_t := G'_t$. In the language of random graphs, such a procedure is also known as the *simple switching* method developed by McKay and Wormald and many others (see, for example, the survey [47] and the references therein).



Two main questions concerning random bipartite graph with prescribed degrees are: what is their number and how long does the Diaconis–Gangolli algorithm take to produce such a random bipartite graph?

In this paper we consider a more specialized version of the Diaconis–Gangolli algorithm. To start with a simple example, on $\mathbb{Z}/2\mathbb{Z}$, consider the problem of generating a random bipartite graph in which every vertex has an even

degree. At time $t = 0$, we start with the empty graph G_0 . At time t , the Diaconis–Gangolli algorithm suggests that from the graph G_{t-1} , choose two pairs of vertices $v_i \neq v_j$ in V and $u_k \neq u_l$ in U uniformly at random. Then, replace the subgraph induced by restricting G_{t-1} on these four vertices by its complement.



Our main theorem below asserts that at time $t \geq \frac{n^2}{8} \log n + 10cn^2 \sqrt{\log n} \log \log n$, the (random) graph G_t is distributed almost uniformly: the L^1 distance between the distribution of G_t and the uniform distribution on the bipartite graphs whose all degrees are even is at most e^{-c} .

Similarly, one can use the same algorithm to generate a uniform bipartite graph with prescribed degree parities, say, the vertices in $A \subset V$ and $B \subset U$ have odd degrees and the rest have even degrees. The only difference is in choosing the starting point G_0 . Instead of starting at the empty graph, one can start at any graph that satisfies the prescribed degree parities.

Associate each bipartite graph with its adjacency matrix A , which is an $n \times n$ matrix whose (i, j) entry is 1 if (v_i, u_j) is an edge and 0 otherwise. The requirement that each vertex has a specific degree parity means that we require A to have given row sums and column sums in $\mathbb{Z}/2\mathbb{Z}$.

More generally, we consider matrices with prescribed row and column sums in $\mathbb{Z}/q\mathbb{Z}$ for positive integers q .

Definition 1.1. A contingency table over $\mathbb{Z}/q\mathbb{Z}$ is an $n \times n$ matrix with entries in $\mathbb{Z}/q\mathbb{Z}$, with fixed row sums $(r_1, \dots, r_n) \in (\mathbb{Z}/q\mathbb{Z})^n$ and fixed column sums $(c_1, \dots, c_n) \in (\mathbb{Z}/q\mathbb{Z})^n$. Note that for the existence of such matrices, we need that $\sum_{i=1}^n r_i = \sum_{j=1}^n c_j$.

The Diaconis–Gangolli random walk on $n \times n$ contingency tables over $\mathbb{Z}/q\mathbb{Z}$ goes as follows. Pick two distinct rows and two distinct columns at random. Then, we look at the equilateral that these rows and columns form. We flip a fair coin. If heads we add

$$T = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

to the corners of the equilateral. If tails, then we add $-T$ to the corners of the equilateral.

Let Ω be the set of all contingency tables with the prescribed degrees in $\mathbb{Z}/q\mathbb{Z}$ and let A_t be the contingency table after t steps of the process. For $A, B \in \Omega$, let $P_A^t(B)$ be the probability of moving from A to B after t steps. The sequence of probability measures P_A^t converges to the uniform measure π on Ω , as $t \rightarrow \infty$, with respect to the total variation distance

$$d(t) := \max_{A \in \Omega} \|P_A^t - \pi\|_{\text{T.V.}} = \max_{A \in \Omega} \left\{ \frac{1}{2} \sum_{B \in \Omega} |P_A^t(B) - \pi(B)| \right\}. \tag{1}$$

A question which arises naturally is to determine the rate of convergence to stationarity of the random walk, which is quantified by the mixing time

$$t_{\text{mix}}(\varepsilon) = \min \{ t \in \mathbb{N} : d(t) \leq \varepsilon \}. \tag{2}$$

Our main result is the following.

Theorem 1.2. *Let $n \geq 4$ and $q \geq 2$. For the Diaconis–Gangolli walk on $n \times n$ contingency tables with entries in $\mathbb{Z}/q\mathbb{Z}$, with row sums $(r_1, \dots, r_n) \bmod q$ and column sums $(c_1, \dots, c_n) \bmod q$, we have that*

(a) (Upper bound) *If $t \geq \frac{n^2}{4(1-\cos \frac{2\pi}{q})} \log n + \frac{n^2}{(1-\cos \frac{2\pi}{q})} 100c \log \log(16n) \sqrt{\log n} \log q$, then*

$$d(t) \leq q^{-c},$$

$$\text{for all } c \geq \frac{640}{\log \log(16n)}.$$

(b) (Lower bound) If $t \leq \frac{n^2}{4(1-\cos \frac{2\pi}{q})} \log n - \frac{n^2}{4(1-\cos \frac{2\pi}{q})} (c + 12) \log q$, then

$$d(t) \geq 1 - q^{-c},$$

for all $c \geq 0$.

Set $t_{n,q} = \frac{n^2}{4(1-\cos \frac{2\pi}{q})} \log n$ and $\Delta_{n,q} = \frac{n^2}{(1-\cos \frac{2\pi}{q})} \log \log(16n) \sqrt{\log n} \log q$. Theorem 1.2 says that over $\mathbb{Z}/q\mathbb{Z}$, the random walk mixes at time $t_{n,q}$ with window of order at most $\Delta_{n,q}$, i.e.

$$\lim_{c \rightarrow \infty} \lim_{n \rightarrow \infty} d(t_{n,q} - c \Delta_{n,q}) = 1^1$$

and

$$\lim_{c \rightarrow \infty} \lim_{n \rightarrow \infty} d(t_{n,q} + c \Delta_{n,q}) = 0.$$

If $\log q = o(\frac{\sqrt{\log n}}{\log \log n})$ then $\Delta_{n,q} = o(t_{n,q})$. In other words, the random walk exhibits *cutoff*, a sharp transition from $d(t) \approx 1$ to $d(t) \approx 0$.

If $n = 2$, then the lazy random walk on contingency tables over $\mathbb{Z}/q\mathbb{Z}$ is the same as the lazy random walk on $\mathbb{Z}/q\mathbb{Z}$, which is known to not have cutoff [27].

To prove the upper bound of Theorem 1.2, we study the density of the spectrum of the transition matrix obtained in Section 2, which we reduce to estimating the number of matrices with a prescribed number of “non-zero” boxes. A “non-zero” box is a submatrix whose entries do not add to zero. One of our biggest difficulties is that the number of non-zero boxes of a matrix can vary significantly by changing the matrix slightly which makes it hard to characterize the matrices with a given number of nonzero boxes. To deal with this difficulty, we provide a way to control this number by the number of nonzero intervals of the matrix rows through a rather simple construction (Lemma 3.6). This then reduces the problem to characterizing matrices with a given number of nonzero intervals of the rows (Lemma 3.7). Since Lemma 3.7 is rather technical, we start by studying an easier case when the number of nonzero boxes is small (Lemma 3.5) to make our general intuition clearer to the reader.

The paper is organized as follows. In Section 1.2, we provide the history of the problem and some remarkable results on related problems. In Section 1.3, we show that the popular probabilistic technique of Markovian couplings is not strong enough to prove Theorem 1.2. In Section 2, we reduce the problem to a random walk on a group and find the spectrum of this random walk using representation theory. Section 3.1 uses Wilson’s method to prove the lower bound of Theorem 1.2. Finally, Section 3.2, which is the heart of our proof, proves the upper bound of Theorem 1.2.

1.2. Literature

Contingency tables are used in statistics, in order to display the results of tests and surveys. Diaconis and Efron ([28, 29]) developed the conditional volume test, which is a method for performing tests of great importance in such tables. The Diaconis–Efron test provides strong motivation for sampling a contingency table with given row and column sums uniformly at random.

Diaconis, Gangolli [30] and Diaconis and Saloff-Coste (see page 373 of [34]) were the first ones to introduce Markov chains for sampling contingency tables, although it is mentioned in [30] that this chain has been used by practitioners. Diaconis and Saloff-Coste proved that if $N = \sum c_i = \sum r_i$ and the number of rows and columns is fixed then the mixing time is of order N^2 . Hernek [38] considered the case when the table has two rows and proved that the same chain mixes in time polynomial in the number of columns and N . Chung, Graham, and Yau [22] proved that a modified version of the Diaconis and Saloff-Coste chain converges in time polynomial in N , the number of rows, and the number of columns, given that N is large.

The Diaconis–Gangolli Markov chain is closely related to card shuffling. In particular, the case of 1-regular bipartite graphs corresponds to the random transposition card shuffle. The famous work of Diaconis and Shahshahani [33] showed that the random transposition walk exhibits cutoff at time $\frac{1}{2}n \log n$.

Dyer, Kannan, and Mount [36] found a new way to sample contingency tables using polytopes which provided the first truly polynomial-time algorithm (polynomial in the number of rows, the number of columns, and the logarithm of N). Later, Morris [44] refined their answers.

¹The lower bound of Theorem 1.2 is in fact stronger than this.

Dyer and Greenhill [35] applied coupling to get pre-cutoff at $n^2 \log N$ for the case of 2×2 heat-bath chain, a different Markov chain on $2 \times n$ contingency tables. Matsui, Matsui and Ono [40] extended the result of [35] for $2 \times 2 \times \dots \times 2 \times J$ contingency tables. Cryan, Dyer, Goldberg, Jerrum, Martin [24] extended the result of [35] for the case where there is a constant number of rows.

A closely related problem to the mixing time is the enumerating problem. It is worth mentioning that there has been a rich literature in enumerating contingency tables ([1–7,9,10,19,25,37,45]) and studying algorithms to sample contingency tables or approximate their number ([12–15,17,20,21,23,34]). On a different note, Blanchet and Stauffer [16] provide a necessary and sufficient condition so that the configuration model outputs a binary contingency table with probability bounded away from 0 and $N \rightarrow \infty$.

Enumerating the graphs with a given degree sequence also has a vast literature and important applications. See for example, Bender and Canfield [11], Bollobás [18] and Wormald [47]. McKay and Wormald used the method of switchings to obtain in [42] an asymptotic formula for the number of labeled graphs with a given degree sequence in the case where the average degree is $o(\sqrt{n})$. See also [8,41,43] for other ranges of the degrees. In a recent breakthrough, Liebenau and Wormald [39] obtained, among other things, the asymptotic number of d -regular graphs on n vertices for all d .

1.3. *There is no Markovian coupling that could give Theorem 1.2*

Coupling is a powerful technique to achieve upper bounds on the mixing time. This section is dedicated to proving that there is no Markovian coupling that would give optimal mixing time bounds.

Proposition 1.3. *Let $q \geq 2$ and let T be a Markovian coupling time for the Diaconis–Gangolli walk in Theorem 1.2, then*

$$\mathbb{P}\left(T > \frac{n^3}{100}\right) \geq \frac{1}{2}.$$

Proof. Each move of the walk is performed with probability $\frac{1}{2\binom{n}{2}}$ for $q \geq 3$ and $\frac{1}{\binom{n}{2}}$ for $q = 2$. At each step we change four entries of the matrix. If there are two copies of the Markov chain $(C_t^1), (C_t^2)$, then we claim that

$$\mathbb{P}(C_t^1 = C_t^2 | C_{t-1}^1 \neq C_{t-1}^2) \leq \frac{6n}{\binom{n}{2}}, \tag{3}$$

for every t . Notice that if $C_{t-1}^1 \neq C_{t-1}^2$ then because of the degree restrictions, C_{t-1}^1 and C_{t-1}^2 have to differ in at least 4 coordinates.

If $C_{t-1}^1 \neq C_{t-1}^2$ and C_{t-1}^1, C_{t-1}^2 differ in at least 9 entries, then (3) holds because the left hand side of (3) is zero. If C_{t-1}^1, C_{t-1}^2 differ in at least 5 vertices, the above probability is at most $\frac{6}{\binom{n}{2}}$. If they differ in four entries a, b, c, d (which is the minimum number of entries they can differ by), the only way to resolve these differences, without creating new differences, is if we change a, b in C_{t-1}^1 , and c, d in C_{t-1}^2 (or any other permutation of these letters). The other side of the box should be the same on the two chains so that we don't create new mismatched coordinates. The total probability of doing such a move is at most $\frac{\binom{4}{2}n}{\binom{n}{2}^2}$, which gives the right hand side term.

We will couple T with a geometric random variable R with probability of success $p = 6n/\binom{n}{2}^2$ so that $T \geq R$ always. We have

$$\mathbb{P}\left(R \geq \frac{1}{3p}\right) = \sum_{k \geq \frac{1}{3p}} (1-p)^{k-1} p \geq (1-p)^{1/(3p)} \geq \frac{1}{2}.$$

Thus, $\mathbb{P}(T \geq \frac{1}{3p}) \geq \mathbb{P}(R \geq \frac{1}{3p}) \geq \frac{1}{2}$ completing the proof of the statement. □

If T is a coupling time then

$$d(t) \leq \mathbb{P}(T > t),$$

and, therefore Proposition 1.3 says that we cannot hope for a Markovian coupling that will give the upper bound of Theorem 1.2 even for $q = 2$.

2. The contingency table walk over $\mathbb{Z}/q\mathbb{Z}$ as a random walk on a group

In this section, we explain how linear algebra and representation theory can be used to prove Theorem 1.2.

Let $A_{i,j,k,l}$ denote the $n \times n$ matrix that has ones on the $(i, k), (j, l)$ positions, -1 on the $(i, l), (j, k)$ positions and zeros else where. If at time t the Markov chain is at a contingency table B_t , then we choose matrices of the form $A_{i,j,k,l}$ and we add them to B_t . Let A_{t+1} be the matrix we choose to add to B_t at time $t + 1$. Then, we have that

$$B_{t+1} = A_{t+1} + B_t = B_0 + A_1 + \dots + A_{t+1}.$$

So, instead of studying the Markov chain (B_t) , we can equivalently study the process C_t , where

$$C_t := A_1 + \dots + A_{t+1}.$$

The advantage of studying C_t is that C_t is a random walk on the group G , which consists of $n \times n$ contingency tables with entries over $\mathbb{Z}/q\mathbb{Z}$ and row sums and column sums zero. This is summarized more formally in the following lemma.

Lemma 2.1. *For the processes (B_t) and (C_t) , we have that*

$$\|\mathbb{P}_x(B_t \in \cdot) - \pi\|_{\text{T.V.}} = \|\mathbb{P}_0(C_t \in \cdot) - U\|_{\text{T.V.}},$$

where $x \in \Omega$, π is the uniform measure on Ω and U is the uniform measure on G .

From now on, we will study the random walk C_t on G . The next lemma characterizes G .

Lemma 2.2. *The group G of $n \times n$ contingency tables with entries over $\mathbb{Z}/q\mathbb{Z}$ and row sums and column sums zero, satisfies $G \sim (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$.*

Proof. G is a vector subspace of $(\mathbb{Z}/q\mathbb{Z})^{n \times n}$ over $\mathbb{Z}/q\mathbb{Z}$. In fact, $G \sim (\mathbb{Z}/q\mathbb{Z})^{n \times n} / A$ where A is the subgroup of $(\mathbb{Z}/q\mathbb{Z})^{n \times n}$ generated by $2n - 1$ relations, that are setting $2n - 1$ rows and columns to be equal to zero. Therefore, $\dim_{\mathbb{Z}/q\mathbb{Z}} G = (n - 1)^2$ and this finishes the proof. \square

Let $B_{i,j}$ be the matrix that has ones in positions $(i, j), (i + 1, j + 1)$, -1 in positions $(i, j + 1), (i + 1, j)$ and zero everywhere else. To diagonalize the matrix P we will need the fact that $\{B_{i,j}\}_{i,j=1}^{n-1}$ is a basis for G . We can see that they are a basis, because they are linearly independent and there are $(n - 1)^2$ of them.

Definition 2.3. Let $\tilde{A}_{i,j,k,l} \in (\mathbb{Z}/q\mathbb{Z})^{(n-1) \times (n-1)}$ be the matrix that has ones on all positions (a, c) that satisfy $i \leq a \leq j$ and $k \leq c \leq l$ and everywhere else zero.

Since

$$A_{i,j,k,l} = \sum_{a=i}^{j-1} \sum_{c=k}^{l-1} B_{a,c},$$

the matrix $\tilde{A}_{i,j-1,k,l-1}$ is the coordinates of $A_{i,j,k,l}$ with respect to the basis $(B_{i,j})$ of G . Similarly, we associate each element of G with its coordinates with respect to the basis (B_{ij}) .

2.1. Fourier transform and the ℓ^2 bound

Let X be a finite group and let S be a symmetric set of generators.

Definition 2.4. Let P be the uniform measure on S . Let ρ be a representation of X . Define the Fourier transform of ρ with respect to P to be

$$\hat{P}(\rho) = \sum_{s \in S} P(s)\rho(s).$$

Theorem 6 of [26, Chapter 3E] says that the Fourier transforms of the irreducible representations of G with respect to P give all of the eigenvalues of P , each one appearing with multiplicity being the dimension of the corresponding representation. The following lemma explains how to use the irreducible representations of X to bound the mixing time of the Markov chain generated by P . It was first used in [31] and the rigorous proof can be found in [26, Chapter 3].

Lemma 2.5 (Upper bound lemma). *For the random walk on X generated by P , we have that*

$$4\|P_{\text{id}}^t - \pi\|_{\text{T.V.}}^2 \leq \sum^* d_\rho \text{Tr}(\hat{P}(\rho)\hat{P}^*(\rho))^t, \tag{4}$$

where d_ρ is the dimension of a representation ρ and the sum is over all irreducible representations ρ of X , but the trivial one.

Apply these results to $X = G$ and S being the set of all $\pm A_{i,j,k,l}$; or equivalently, $X = (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$ and S being the set of all $\pm \tilde{A}_{i,j,k,l}$. Let $y, g \in (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$, define

$$\rho_y(g) = e^{\frac{2\pi i \langle g, y \rangle}{q}},$$

where $\langle g, y \rangle = \sum_{i=1}^{(n-1)^2} g_i y_i$ is the inner product of y, g . Each ρ_y is one dimensional, therefore it is irreducible. As explained in Lemma 2 of [26], we have that

$$\sum d_i^2 = q^{(n-1)^2},$$

where the sum is taken over all irreducible representations and d_i is the dimension of each irreducible representation. Therefore, the set $\{\rho_y, y \in (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}\}$ consists of all irreducible representations of G .

The following proposition computes the Fourier transform of each irreducible ρ_y with respect to P .

Lemma 2.6. *Let $y \in (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$, then*

$$\hat{P}(\rho_y) = \frac{1}{\binom{n}{2}} \sum_{i,j,k,l} \cos \frac{2\pi \langle y, \tilde{A}_{i,j,k,l} \rangle}{q},$$

where $\tilde{A}_{i,j,k,l}$ is defined in Definition 2.3.

Proof. The proof of the lemma follows from the definition of ρ_y and the fact that

$$\exp\left(\frac{2\pi i \langle \tilde{A}_{i,j,k,l}, y \rangle}{q}\right) + \exp\left(\frac{-2\pi i \langle \tilde{A}_{i,j,k,l}, y \rangle}{q}\right) = 2 \cos \frac{2\pi \langle y, \tilde{A}_{i,j,k,l} \rangle}{q}. \quad \square$$

Applying Lemmas 2.5 and 2.6, we obtain the following bound for the random walk (C_t) on G and the uniform measure U on G as in Lemma 2.1.

$$\|P_0(C_t \in \cdot) - U\|_{\text{T.V.}}^2 \leq \sum_{y \in G \setminus \{0\}} \hat{P}(\rho_y)^{2t} = \sum_{y \in G \setminus \{0\}} \left(\frac{1}{\binom{n}{2}} \sum_{i,j,k,l} \cos \frac{2\pi \langle y, \tilde{A}_{i,j,k,l} \rangle}{q} \right)^{2t}. \tag{5}$$

2.2. Bounding negative eigenvalues

In this section, we show that the negative eigenvalues are bounded away from minus one and, therefore, the don't contribute much to the right hand side of (4).

Let $S = \{\pm A_{i,j,k,l} : 1 \leq i < j \leq n, 1 \leq k < l \leq n\}$ denote the set of generators of the Diaconis–Gangolli random walk of interest over $\mathbb{Z}/q\mathbb{Z}$. Let P be the transition matrix and $Q = \frac{1}{2}(I + P)$ where I is the identity matrix. Notice that since all the eigenvalues $\hat{P}(\rho_y)$ of P lie in $[-1, 1]$, we have that all the eigenvalues of Q are non-negative, real numbers.

Let $f : G \rightarrow \mathbb{R}$ be a function and let

$$\mathcal{F}(f, f) = \sum_{x,y \in G} (f(x) + f(y))^2 P(x, y) \quad \text{and} \quad \tilde{\mathcal{F}}(f, f) = \sum_{x,y \in G} (f(x) + f(y))^2 Q(x, y). \tag{6}$$

We are going to use Lemma 4 of [31] to bound the negative eigenvalues of P from below. For completeness, we rewrite the statement of Lemma 4 of [31] for our case.

Lemma 2.7 ([31], Lemma 4). *If $\tilde{\mathcal{F}}(f, f) \leq A_* \mathcal{F}(f, f)$ for every $f : G \rightarrow \mathbb{R}$, then*

$$\hat{P}(\rho_y) \geq -1 + \frac{1}{A_*},$$

for every $y \in (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$.

We are going to use Lemma 2.7 to prove the following bound on the eigenvalues of P .

Lemma 2.8. *For the Diaconis–Gangolli random walk on contingency tables over $\mathbb{Z}/q\mathbb{Z}$, we have that*

$$\hat{P}(\rho_y) \geq -\frac{54}{55},$$

for every $y \in (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$ and $n \geq 4$.

Proof. The proof is based on the technique of “flows” as presented in Theorem 2.3 of [32]. For completion, we explain how this method works.

Notice that for every $f : G \rightarrow \mathbb{R}$, we have that

$$\begin{aligned} \tilde{\mathcal{F}}(f, f) &= 2 \sum_{x \in G} f(x)^2 + \frac{1}{2} \sum_{x \neq y \in G} (f(x) + f(y))^2 P(x, y) \\ &= 2 \sum_{x \in G} f(x)^2 + \frac{1}{2} \mathcal{F}(f, f). \end{aligned} \tag{7}$$

To bound the term $\sum_{x \in G} f(x)^2$, we observe that for $n \geq 3$ and for any $k < b < l$ and $i < j$,

$$A_{i,j,k,l} = A_{i,j,k,b} + A_{i,j,b,l}, \tag{8}$$

which can be illustrated for the case $i = 1 = k$, $j = 2 = b$ and $l = 3$ as

$$\begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \\ \dots & & \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ \dots & & \end{bmatrix} + \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \\ \dots & & \end{bmatrix}.$$

Let $x \in G$ and let $f : G \rightarrow \mathbb{R}$. Using equation (8), we have

$$\begin{aligned} 2f(x) &= [f(x) + f(x + A_{i,j,k,b})] - [f(x + A_{i,j,k,b}) + f(x + A_{i,j,k,b} + A_{i,j,b,l})] \\ &\quad + [f(x + A_{i,j,k,l}) + f(x)]. \end{aligned}$$

Applying Cauchy–Schwartz gives

$$\begin{aligned} \frac{4}{3} f(x)^2 &\leq [f(x) + f(x + A_{i,j,k,b})]^2 + [f(x + A_{i,j,k,b}) + f(x + A_{i,j,k,b} + A_{i,j,b,l})]^2 \\ &\quad + [f(x + A_{i,j,k,l}) + f(x)]^2. \end{aligned}$$

Thus, by averaging,

$$\begin{aligned} f(x)^2 &\leq \frac{3}{4 \binom{n}{2} \binom{n}{3}} \sum_{i < j; k < b < l} \{ [f(x) + f(x + A_{i,j,k,b})]^2 + [f(x + A_{i,j,k,b}) + f(x + A_{i,j,k,b} + A_{i,j,b,l})]^2 \\ &\quad + [f(x + A_{i,j,k,l}) + f(x)]^2 \}. \end{aligned}$$

Using the identities $P(x, x + g) = \frac{1}{|S|} \mathbf{1}_{g \in S}$, $|S| = \binom{n}{2}$ if $q = 2$ and $|S| = 2\binom{n}{2}$ if $q > 2$, we get

$$\begin{aligned} \sum_{x \in G} f(x)^2 &\leq \frac{3}{4\binom{n}{2}\binom{n}{3}} \sum_{x \in G} \sum_{i < j; k < b < l} \{ [f(x) + f(x + A_{i,j,k,b})]^2 \\ &\quad + [f(x + A_{i,j,k,b}) + f(x + A_{i,j,k,b} + A_{i,j,b,l})]^2 + [f(x + A_{i,j,k,l}) + f(x)]^2 \} \\ &\leq \frac{9n}{4\binom{n}{2}\binom{n}{3}} \sum_{x \in G, g \in S} (f(x) + f(x + g))^2 \leq \frac{18n\binom{n}{2}}{4\binom{n}{3}} \sum_{x \in G, g \in S} (f(x) + f(x + g))^2 P(x, x + g) \\ &\leq 27 \sum_{x \in G, g \in S} (f(x) + f(x + g))^2 P(x, x + g). \end{aligned} \tag{9}$$

Equation (7) and (9) give that

$$\tilde{\mathcal{F}}(f, f) \leq 54\mathcal{F}(f, f) + \frac{1}{2}\mathcal{F}(f, f) \leq 55\mathcal{F}(f, f)$$

and then, Lemma 2.7 for $A_* = 55$ gives that

$$\hat{P}(\rho_y) \geq -\frac{54}{55},$$

for every $y \in (\mathbb{Z}/q\mathbb{Z})^{(n-1)^2}$. □

3. Proof of Theorem 1.2

3.1. Proof of the lower bound

For the proof of the lower bound, we will use Wilson’s lemma.

Lemma 3.1 (Lemma 5, [46]). *Let ε, R be positive numbers and $0 < \gamma < 2 - \sqrt{2}$. Let $F : X \rightarrow \mathbb{R}$ be a function on the state space X of a Markov chain (C_t) such that*

$$\mathbb{E}[F(C_{t+1})|C_t] = (1 - \gamma)F(C_t), \quad \mathbb{E}[[F(C_{t+1}) - F(C_t)]^2|C_t] \leq R,$$

and

$$t \leq \frac{\log \max_{x \in X} F(x) + \frac{1}{2} \log(\gamma\varepsilon/(4R))}{-\log(1 - \gamma)}.$$

Then the total variation distance from stationarity at time t is at least $1 - \varepsilon$.

Proof of the lower bound of Theorem 1.2. As in Definition 2.3 and the discussion that follows, we represent each element of G as an $(n - 1) \times (n - 1)$ matrix which is its coordinates with respect to the basis (B_{ij}) of G .

Let $D_{a,b}$ be the $(n - 1) \times (n - 1)$ matrix with entries at $(2a - 1, 2b - 1), (2a, 2b)$, equal to one, entries $(2a, 2b - 1), (2a - 1, 2b)$ equal to -1 , while all the other entries are zero. Theorem 6 of Chapter 3E of [26] says that the functions $G_{a,b}(x) = \cos(\frac{2\pi(x, D_{a,b})}{q})$ are eigenfunctions of the transition matrix P . To see this, consider

$$D_{1,1} = \begin{pmatrix} 1 & -1 & 0 & \dots \\ -1 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

and let $C \in (\mathbb{Z}/q\mathbb{Z})^{(n-1) \times (n-1)}$. We have that

$$G_{1,1}(C) = \cos\left(\frac{2\pi(C(1, 1) - C(1, 2) - C(2, 1) + C(2, 2))}{q}\right).$$

We want to study how $G_{1,1}(C_t)$ changes after one step. Each move of the walk corresponds to choosing a sub-matrix inside C_t and adding 1 to all its entries with probability $1/2$ or subtracting 1 from all of its entries with probability $1/2$. The only moves that affect the value of $G_{1,1}(C_t)$ are the ones that affect exactly one of the following entries $C_t(1, 1)$, $C_t(1, 2)$, $C_t(2, 1)$ and $C_t(2, 2)$. There are $2(n - 1)^2$ such moves. Given that each move happens with probability $\frac{1}{2\binom{n}{2}}$,

we have that

$$\begin{aligned} \mathbb{E}[G_{1,1}(C_{t+1})|C_t] &= \left(1 - \frac{4}{n^2}\right) \cos\left(\frac{2\pi(C_t(1, 1) - C_t(1, 2) - C_t(2, 1) + C_t(2, 2))}{q}\right) \\ &\quad + \frac{4}{2n^2} \cos\left(\frac{2\pi(C_t(1, 1) - C_t(1, 2) - C_t(2, 1) + C_t(2, 2) + 1)}{q}\right) \\ &\quad + \frac{4}{2n^2} \cos\left(\frac{2\pi(C_t(1, 1) - C_t(1, 2) - C_t(2, 1) + C_t(2, 2) - 1)}{q}\right) \\ &= \left(1 - \frac{4}{n^2} + \frac{4}{n^2} \cos\left(\frac{2\pi}{q}\right)\right) G_{1,1}(C_t), \end{aligned} \tag{10}$$

where in (10) we used the identity $\cos(a + b) + \cos(a - b) = 2 \cos a \cos b$. Similarly, for any $G_{a,b}$, we have that

$$\mathbb{E}[G_{a,b}(C_{t+1})|C_t] = \left(1 - \frac{4}{n^2} + \frac{4}{n^2} \cos\left(\frac{2\pi}{q}\right)\right) G_{a,b}(C_t). \tag{11}$$

To apply Wilson’s lemma, we want to consider an eigenfunction that has a big maximum value but a reasonably small variance. Therefore, we consider the function

$$F(x) = \sum_{a,b=0}^{\lfloor \frac{n-1}{2} \rfloor} G_{a,b}.$$

Then, we have that $\max F = F(0) = \lfloor \frac{n-1}{2} \rfloor^2$ and (11) gives that

$$\mathbb{E}[F(C_{t+1})|C_t] = \left(1 - \frac{4}{n^2} + \frac{4}{n^2} \cos\left(\frac{2\pi}{q}\right)\right) F(C_t). \tag{12}$$

Finally, we have that

$$\mathbb{E}[(F(C_{t+1}) - F(C_t))^2|C_t] \leq 64. \tag{13}$$

This is because every move $\tilde{A}_{i,j,k,l}$ that we might choose to make, affects the $G_{a,b}$ who share a unique one with $\tilde{A}_{i,j,k,l}$. For example,

$$\tilde{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & \dots \\ 1 & 1 & 1 & 0 & \dots \\ 1 & 1 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \dots & & & & \dots \end{pmatrix}$$

fixes all $G_{a,b}$, but $G_{2,2}$. Therefore, every move that we make can affect at most four $G_{a,b}$. Thus, $|F(C_{t+1}) - F(C_t)| \leq 8$ surely and so $\mathbb{E}[(F(C_{t+1}) - F(C_t))^2|C_t] \leq 64$.

Using equations (11) and (13) for $n \geq 4$, we have that $\gamma = \frac{4}{n^2}(1 - \cos \frac{2\pi}{q}) \leq \frac{1}{2} < 2 - \sqrt{2}$ and $R = 64$. Since $\frac{1}{-\log(1-\gamma)} \geq \frac{1}{\gamma} - 1$, Lemma 3.1 asserts that if

$$t \leq \left(\frac{1}{\gamma} - 1\right) \log(F(0)\gamma^{1/2}\varepsilon^{1/2}R^{-1/2}/2)$$

then

$$d(t) \geq 1 - \varepsilon.$$

Writing $\varepsilon = q^{-c}$, we get that

$$\begin{aligned} \left(\frac{1}{\gamma} - 1\right) \log(F(0)\gamma^{1/2}\varepsilon^{1/2}R^{-1/2}/2) &\geq \left(\frac{n^2}{4(1 - \cos \frac{2\pi}{q})} - 1\right) \log\left(\frac{n\sqrt{1 - \cos \frac{2\pi}{q}}}{72q^{c/2}}\right) \\ &\geq \frac{n^2}{4(1 - \cos \frac{2\pi}{q})} \log n - \frac{n^2 \log q}{4(1 - \cos \frac{2\pi}{q})} (c + 12), \end{aligned}$$

where we used $n \geq 4$, $q \geq 2$, and $1 - \cos \frac{2\pi}{q} \leq 2$. Thus, if

$$t \leq \frac{n^2}{4(1 - \cos \frac{2\pi}{q})} \log n - \frac{n^2 \log q}{4(1 - \cos \frac{2\pi}{q})} (c + 12),$$

then $d(t) \geq 1 - q^{-c}$ as desired. □

3.2. Proof of the upper bound of Theorem 1.2

By (1) and Lemma 2.1, we get

$$d(t)^2 \leq \|\mathbb{P}_0(C_t \in \cdot) - U\|_{T,V}^2,$$

where C_t and U are as in Lemma 2.1. Combining this with (5), we get

$$d(t)^2 \leq \sum_y \hat{P}(\rho_y)^{2t}, \tag{14}$$

where the sum runs over all nonzero $(n - 1) \times (n - 1)$ matrices y and

$$\hat{P}(\rho_y) = \frac{1}{\binom{n}{2}} \sum_{i,j,k,l} \cos \frac{2\pi \langle y, \tilde{A}_{i,j,k,l} \rangle}{q}.$$

For each $a \in \mathbb{Z}/q\mathbb{Z}$, let $N_a(y)$ be the number of quadruples (i, j, k, l) with $1 \leq i \leq j \leq n - 1$ and $1 \leq k \leq l \leq n - 1$ such that $\langle y, \tilde{A}_{i,j,k,l} \rangle = a$. Then,

$$\hat{P}(\rho_y) = \frac{1}{\binom{n}{2}} \sum_{a \in \mathbb{Z}/q\mathbb{Z}} N_a(y) \cos \frac{2\pi a}{q}. \tag{15}$$

Note that $\sum_{a \neq 0} N_a(y) = \binom{n}{2}^2 - N_0(y)$.

We need to show that for $t = \frac{n^2}{4(1 - \cos \frac{2\pi}{q})} \log n + \frac{n^2}{(1 - \cos \frac{2\pi}{q})} 100c \log \log(16n) \sqrt{\log n} \log q$,

$$d(t)^2 \leq q^{-2c}.$$

To do so, we decompose the right-hand side of (14) into the sum over all y with $\hat{P}(\rho_y) < 0$ and the sum of the rest, and denote the corresponding sums by d_1 and d_2 .

By Lemma 2.8, each of the $\hat{P}(\rho_y)$ with $\hat{P}(\rho_y) < 0$ satisfies $\hat{P}(\rho_y) \geq -\frac{54}{55}$. Thus,

$$d_1 \leq q^{(n-1)^2} \left(\frac{54}{55}\right)^{2t} \leq q^{n^2} q^{-8n^2-2c} \leq q^{-2c}/2, \tag{16}$$

where we used the assumption that $c \geq \frac{640}{\log \log(16n)}$ and $n \geq 4$ to get the estimate

$$t \geq \frac{n^2}{(1 - \cos \frac{2\pi}{q})} 100c \log \log(16n) \sqrt{\log n} \log q \geq c \log_{55/54} q + 60n^2 \log_{55/54} q.$$

For the rest of the proof, we will show that $d_2 \leq q^{-2c}/2$. Assume that it holds, we have $d(t)^2 \leq d_1 + d_2 \leq q^{-2c}$ as desired.

For each $(n - 1) \times (n - 1)$ matrix y and each $1 \leq i \leq j \leq n - 1$ and $1 \leq k \leq l \leq n - 1$, we consider the minor $y_{i,j,k,l}$ obtained by restricting to the rows from i to j and the columns from k to l , inclusively. We say that $[i, j] \times [k, l]$ is a *nonzero box* if the sum of the entries in this minor is nonzero. Let

$$N(y) := \text{the number of non-zero boxes of } y. \tag{17}$$

Observe that $\sum_{a \neq 0} N_a(y) = N(y)$. When $\hat{P}(\rho_y) \geq 0$, we have from (15) that

$$0 \leq \hat{P}(\rho_y) \leq \frac{1}{\binom{n}{2}^2} \left[N_0(y) + \left(\sum_{a \neq 0} N_a(y) \right) \cos \frac{2\pi}{q} \right] = 1 - \frac{(1 - \cos \frac{2\pi}{q})N(y)}{\binom{n}{2}^2}.$$

Thus,

$$d_2 \leq \sum_{y: \hat{P}(\rho_y) \geq 0} \left(1 - \frac{(1 - \cos \frac{2\pi}{q})N(y)}{\binom{n}{2}^2} \right)^{2t} \leq \Sigma,$$

where

$$\Sigma := \sum \left(1 - \frac{(1 - \cos \frac{2\pi}{q})N(y)}{\binom{n}{2}^2} \right)^{2t}, \tag{18}$$

where the sum is taken over all matrices $y \neq 0$ with $N(y) \leq \frac{\binom{n}{2}^2}{1 - \cos \frac{2\pi}{q}}$.

It remains to show that

$$\Sigma \leq q^{-2c}/2 \tag{19}$$

and for that we need to control the number of matrices y with a prescribed range of $N(y)$.

The number $N(y)$ of nonzero boxes of a matrix y can vary significantly just by changing an entry of y . For example, if y is the zero matrix then $N(y) = 0$. If we add a single entry 1 at around the middle of the matrix y , then $N(y) \approx (n/2)^4$ which is significantly larger. Now, if the single entry 1 is at position $(1, 1)$ (for example) instead of at the middle of the matrix, then $N(y) \approx n^2$.

It turns out to be useful to look at one dimensional version of the nonzero boxes. Let u be a vector in $(\mathbb{Z}/q\mathbb{Z})^{n-1}$. For any $1 \leq i \leq j \leq n - 1$, the interval $[i, j]$ is said to be a *nonzero interval* of u if the sum of the entries of the vector $[u(i), u(i + 1), \dots, u(j)]$ is nonzero in $\mathbb{Z}/q\mathbb{Z}$. Let $S(u)$ be the number of nonzero intervals in u .

In order to control $S(u)$, we introduce the following definition.

Definition 3.2. Let the *skeleton* of u to be the set $I(u) = \{i_1, i_2, \dots, i_{s(u)}\} \subset [1, n - 1]$ with

- $i_1 \geq 1$ being the smallest index such that $u_{i_1} \neq 0$,
- i_k being the smallest index such that $i_k \geq i_{k-1} + 2$ and $u_{i_k} \neq 0$, for all $2 \leq k \leq s(u)$,
- $u_i = 0$ for all $i \geq i_{s(u)} + 2$.

For example, if u has nonzero entries at positions 3, 4, 5, 8, 9 and the rest are 0 then $I(u) = \{3, 5, 8\}$ and $s(u) = 3$.

$$\begin{array}{cccccccccccc}
 u = [& 0 & 0 & * & * & * & 0 & 0 & * & * & 0 & 0 & \dots & 0] \\
 & & & & & & & & & & & & & & \\
 & & & & \downarrow & \downarrow & & & \downarrow & & & & & & \\
 I(u) = \{ & & & & 3 & 5 & & & 8 & & & & & & \}
 \end{array}$$

Observe that for any row vector u , the skeleton size $s(u)$ is at most $n/2$ and indices of the nonzero entries of u form a subset of $I(u) \cup (I(u) + 1)$. Thus, the number of nonzero elements in u is at most $2s(u)$.

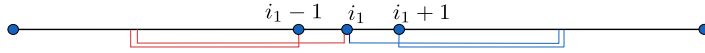
The number of nonzero intervals $S(u)$ is controlled by the skeleton size $s(u)$ as follows.

Lemma 3.3. *Let u be a nonzero $1 \times (n - 1)$ vector in $(\mathbb{Z}/q\mathbb{Z})^{n-1}$. Then for every $n \geq 4$ and $q \geq 2$, we have*

$$S(u) \geq s(u)(n - s(u)).$$

Furthermore, the number of nonzero intervals with one or both endpoints belonging to $I(u) \cup (I(u) \pm 1)$ is at least $s(u)(n - s(u))$.

Proof. Let the skeleton of u be $I(u) = \{i_1, i_2, \dots, i_{s(u)}\}$ with $i_1 < i_2 < \dots < i_{s(u)}$. We first claim that the number of nonzero intervals $[i, j]$ with either $j \in \{i_1 - 1, i_1\}$ or $i \in \{i_1, i_1 + 1\}$ or both² is at least $n - 1$.



Indeed, let $1 \leq t \leq n - 1$ be any index. If $t < i_1$ then either $[t, i_1 - 1]$ or $[t, i_1]$ is a nonzero interval. If $t = i_1$ then $[t, i_1]$ is a nonzero interval. If $t > i_1$ then $[i_1, t]$ or $[i_1 + 1, t]$ is a nonzero interval.

Applying the same argument for i_2 in place of i_1 and t running from 1 to $n - 1$ except for $t = i_1, i_1 + 1$ (to avoid double counting), we obtain at least $n - 3$ other nonzero intervals $[i, j]$ with either $j \in \{i_2 - 1, i_2\}$ or $i \in \{i_2, i_2 + 1\}$ or both. Keep running this argument for $i_3, \dots, i_{s(u)}$ gives

$$(n - 1) + (n - 3) + (n - 5) + \dots + (n - 2s(u) + 1) = ns(u) - s(u)^2$$

nonzero intervals as claimed. □

Using a similar argument, we can control the number nonzero boxes by the number of nonzero intervals.

Lemma 3.4. *Assume that an $(n - 1) \times (n - 1)$ matrix y has some nonzero rows i_1, i_2, \dots, i_k with $|i_m - i_l| \geq 2$ for all $m \neq l$. Let S_m be the number of nonzero intervals of row i_m . Then for every $n \geq 4$ and $q \geq 2$, we have*

$$N(y) \geq S_1(n - 1) + S_2(n - 3) + \dots + S_k(n - 2k + 1).$$

Proof. Consider a nonzero interval $[p, q]$ of row i_m for some $1 \leq p \leq q \leq n - 1$ and $1 \leq m \leq k$. For each $1 \leq r \leq i_m - 1$, either the box $[r, i_m - 1] \times [p, q]$ or the box $[r, i_m] \times [p, q]$ is a nonzero box. Similarly, for each $i_m + 1 \leq r \leq n - 1$, either the box $[i_m + 1, r] \times [p, q]$ or the box $[i_m, r] \times [p, q]$ is a nonzero box. For $r = i_m$, the box $[i_m, r] \times [p, q]$ is itself a nonzero box. Thus, each nonzero interval $[p, q]$ contributes at least $n - 1$ nonzero boxes that touch³ the columns p, q and either touch the row i_m or the row $i_m - 1$ from above or the row $i_m + 1$ from below. Taking union over all nonzero intervals of row i_m , there are at least $S_m(n - 1)$ nonzero boxes that either touch the row i_m or the row $i_m - 1$ from above or the row $i_m + 1$ from below. Taking union over m and subtracting the multiple-counted boxes, we conclude that the number of nonzero boxes is at least $S_1(n - 1) + S_2(n - 3) + \dots + S_k(n - 2k + 1)$. □

Now, we give an upper bound on the number of matrices y with $o(n^3)$ nonzero boxes.

Lemma 3.5. *Let $0 < \varepsilon_n \leq \frac{1}{80}$ be any number. Let w be a positive integer satisfying $1 \leq w \leq \varepsilon_n n$. Then for every $n \geq 4$ and $q \geq 2$, the number of $(n - 1) \times (n - 1)$ matrices y in $\mathbb{Z}/q\mathbb{Z}$ with*

$$N(y) \leq \left(w + \frac{1}{2}\right)n^2 \tag{20}$$

is at most $q^{32w} n^{2w+1+60\varepsilon_n w}$.

Moreover, if $w \leq \min\{\frac{1}{100\varepsilon_n}, \varepsilon_n n\}$, the number of such matrices y is at most $q^{32w} n^{2w}$. Also, for such w , the number of matrices y with

$$N(y) \leq (w - 40w\varepsilon_n)n^2 \tag{21}$$

is at most $q^{16w} n^{2(w-1)}$.

²This guarantees that one or both endpoints of $[i, j]$ belong to the set $\{i_1, i_1 \pm 1\}$.

³We say that a box touches a row or a column if it has that row or column as a boundary. In other words, the box $[i, j] \times [k, l]$ touches rows i, j and columns k, l .

In the above statement, only the constant 2 in n^{2w} is important. All of the other constants 16, 32, 80, 100 are merely for explicitness. Their exact values do not play any significant role. This also holds for other (big) constants in the rest of the proof.

Proof. To prove the first part of the lemma, let y be an $(n - 1) \times (n - 1)$ matrix satisfying (20). Let $i_1 < i_2 < \dots < i_p$ be some nonzero rows of y with $i_m \geq i_{m-1} + 2$ for each m . For each row i , let S_i be the number of nonzero intervals in that row and $0 \leq s_i \leq \frac{n}{2}$ be the size of the its skeleton. By Lemma 3.3, $S_i \geq s_i(n - s_i)$. Note that $s_{i_j} \geq 1$ for all $j = 1, \dots, p$ as they are nonzero rows. By applying Lemma 3.4 to the row i_j , we get

$$\left(w + \frac{1}{2}\right)n^2 \geq S_{i_j}(n - 1),$$

which together with Lemma 3.3 give

$$\left(w + \frac{1}{2}\right)n^2 \geq s_{i_j}(n - s_{i_j})(n - 1).$$

Thus, for each $j = 1, \dots, p$,

$$s_{i_j} \leq 2s_{i_j} \left(1 - \frac{s_{i_j}}{n}\right) \leq 4w \leq 4\varepsilon_n n. \tag{22}$$

By applying Lemma 3.4 to the rows i_1, \dots, i_p , we get

$$\left(w + \frac{1}{2}\right)n^2 \geq S_{i_1}(n - 1) + S_{i_2}(n - 3) + \dots + S_{i_p}(n - 2p + 1) \tag{23}$$

and by Lemma 3.3,

$$\left(w + \frac{1}{2}\right)n^2 \geq s_{i_1}(n - s_{i_1})(n - 1) + s_{i_2}(n - s_{i_2})(n - 3) + \dots + s_{i_p}(n - s_{i_p})(n - 2p + 1). \tag{24}$$

Combining (22) and (24), we obtain

$$w + \frac{1}{2} \geq \sum_{j=1}^p s_{i_j} \left(1 - \frac{s_{i_j}}{n}\right) \left(1 - \frac{2j - 1}{n}\right) \geq \sum_{j=1}^p s_{i_j} (1 - 4\varepsilon_n) \left(1 - \frac{2j - 1}{n}\right). \tag{25}$$

Thus,

$$w + \frac{1}{2} \geq (1 - 4\varepsilon_n) \sum_{j=1}^p \left(1 - \frac{2j - 1}{n}\right) = (1 - 4\varepsilon_n)p \left(1 - \frac{p}{n}\right) \geq p/4,$$

because $p \leq n/2$ by the assumption that $i_m \geq i_{m-1} + 2$ for $m = 2, \dots, p$. So, $p \leq 4w + 2 \leq 6\varepsilon_n n$. Plugging this into (25), we obtain

$$\left(w + \frac{1}{2}\right)(1 + 20\varepsilon_n) \geq \left(w + \frac{1}{2}\right)(1 - 4\varepsilon_n)^{-1}(1 - 12\varepsilon_n)^{-1} \geq \sum_{j=1}^p s_{i_j} \geq p. \tag{26}$$

We conclude that if there are p nonzero rows whose indices are of distance at least 2 from each other, then $p \leq (w + \frac{1}{2})(1 + 20\varepsilon_n)$. Let

$$f(w) = \left\lfloor \left(w + \frac{1}{2}\right)(1 + 20\varepsilon_n) \right\rfloor.$$

Note that $f(w) \leq 2w \leq n/2$. Thus, there are at most

$$\sum_{p=1}^{f(w)} \binom{n}{p} 2^p \leq f(w) \binom{n}{f(w)} 2^{f(w)}$$

ways to choose the indices of the nonzero rows of y , and it's always true that there are at most $2f(w)$ nonzero rows. Similarly, there are at most that many ways to choose the indices of the nonzero columns of y . Let \mathcal{I} be the set of the chosen columns. We have $|\mathcal{I}| \leq 2f(w)$. Applying (26) to the sequence of chosen rows with odd indices i_1, \dots, i_p , we see that the number of ways to choose the corresponding value s_{i_j} for these rows is at most the number of ways to choose a sequence of positive integers s_{i_1}, \dots, s_{i_p} satisfying (26) for some value p . And that number is at most

$$\sum_{a=1}^{f(w)} \sum_{p=1}^a |\{(s_{i_1}, \dots, s_{i_p} : s_{i_1} + \dots + s_{i_p} = a)\}| = \sum_{a=1}^{f(w)} \sum_{p=1}^a \binom{a-1}{p-1} = \sum_{a=1}^{f(w)} 2^{a-1} \leq 2^{f(w)}.$$

Having chosen the s_i for the rows with odd indices, by the definition of skeletons and the fact that the nonzero entries lie on the columns in \mathcal{I} and (26), the number of ways to choose these rows in $(\mathbb{Z}/q\mathbb{Z})^{n-1}$ is at most

$$\prod_{j=1}^p q^{2s_{i_j}} \binom{|\mathcal{I}|}{s_{i_j}} \leq \prod_{j=1}^p q^{2s_{i_j} |\mathcal{I}|^{s_{i_j}}} \leq q^{2f(w)} (2f(w))^{f(w)}. \tag{27}$$

Similarly for the rows with even indices.

All in all, the number of choices for y is at most

$$\left[f(w) \binom{n}{f(w)} 2^{f(w)} \right]^2 [2^{f(w)}]^2 [q^{2f(w)} (2f(w))^{f(w)}]^2 \leq q^{32w} n^{2f(w)} \leq q^{32w} n^{2w+1+60\varepsilon_n n}.$$

Now, for the second part of the lemma, if $w \leq \frac{1}{100\varepsilon_n}$ then $f(w) \leq w + \frac{5}{6}$. In that case, $f(w) \leq w$ because both are integers. Thus, the number of matrices y satisfying (20) is at most $q^{32w} n^{2w}$ as claimed. To bound the number of matrices y that satisfy (21), we use the same argument, with $w + 1/2$ being replaced by $w - 40w\varepsilon_n$ throughout. We conclude that the number of matrices y satisfying (21) is at most $q^{16w} n^{2f(w)}$ where $f(w)$ is the integer part of

$$(w - 40w\varepsilon_n)(1 + 20\varepsilon_n),$$

which is strictly less than w . Therefore, $f(w) \leq w - 1$. This completes the proof. □

It is more difficult to handle the matrices with a large number of nonzero boxes. We shall need the following stronger version of Lemma 3.4.

Lemma 3.6. *Let y be an $(n - 1) \times (n - 1)$ matrix. Let Ψ_i be a collection of nonzero intervals on each row i such that for every two consecutive rows i and $i + 1$, we have that $\Psi_i \cap \Psi_{i+1} = \emptyset$. Then for every $n \geq 4$ and $q \geq 2$, it holds that*

$$2N(y) \geq (|\Psi_1| + |\Psi_2| + \dots + |\Psi_{n-1}|)n. \tag{28}$$

Proof. For each row i , let \mathfrak{P}_i be the collection of nonzero boxes of the form $[i, j] \times [k, l]$ or $[i + 1, j] \times [k, l]$ where $j \geq i + 1$ and $[k, l] \in \Psi_i$ together with the nonzero boxes $[i, i] \times [k, l]$ (for all $[k, l] \in \Psi_i$). Note that for each $[k, l] \in \Psi_i$, since the interval $[k, l]$ is a nonzero interval in row i , either $[i, j] \times [k, l]$ or $[i + 1, j] \times [k, l]$ must be a nonzero box. Thus, $|\mathfrak{P}_i| \geq |\{j : j \geq i + 1\}| |\Psi_i| + |\Psi_i| = (n - i) |\Psi_i|$. By the hypothesis, the $(\mathfrak{P}_i)_{i=1}^{n-1}$ are disjoint. Thus,

$$N(y) \geq \sum_{i=1}^{n-1} |\mathfrak{P}_i| \geq \sum_{i=1}^{n-1} (n - i) |\Psi_i|. \tag{29}$$

Likewise, let \mathfrak{Q}_j be the collection of nonzero boxes of the form $[j, i] \times [k, l]$ or $[j, i - 1] \times [k, l]$ where $j \leq i - 1$ and $[k, l] \in \Psi_i$ together with the nonzero boxes $[i, i] \times [k, l]$ (for all $[k, l] \in \Psi_i$). Then the \mathfrak{Q}_i are disjoint and $|\mathfrak{Q}_i| \geq i |\Psi_i|$. Therefore,

$$N(y) \geq \sum_{i=1}^{n-1} |\mathfrak{Q}_i| \geq \sum_{i=1}^{n-1} i |\Psi_i|. \tag{30}$$

Adding up (29) and (30), we obtain the desired bound. □

Now, we bound the number of matrices y with a large number of nonzero boxes.

Lemma 3.7. *Let ε_n be a positive number satisfying $\frac{1}{100\sqrt{\log n}} \leq \varepsilon_n \leq \frac{1}{80}$. Let $w \geq \varepsilon_n n$ be an integer. For every $n \geq 4$ and $q \geq 2$, the number of $(n - 1) \times (n - 1)$ matrices y with*

$$N(y) \leq \left(w + \frac{1}{2}\right)n^2 \tag{31}$$

is at most $n^{2w+C_1\varepsilon_n w}$ where $C_1 = \frac{5 \cdot 10^3 \log q}{\varepsilon_n \sqrt{\log n}} + 6 \cdot 10^4$.

Later, we will use this lemma with $\varepsilon_n = \frac{1}{100\sqrt{\log n}}$.

Proof. We will treat the “big” rows and “small” rows separately.

Step 1: Big rows. A row i is said to be a big row if either $S_{i-1} \geq \varepsilon_n n(n - \varepsilon_n n)$ or $S_i \geq \varepsilon_n n(n - \varepsilon_n n)$ or $S_{i+1} \geq \varepsilon_n n(n - \varepsilon_n n)$ where S_j is the number of nonzero interval of row j . In other words, either row i or $i \pm 1$ has a large number of nonzero intervals.

Let \mathcal{B} be the set of indices i of the big rows in y .

Let M be the number of rows i with

$$S_i \geq \varepsilon_n n(n - \varepsilon_n n) \geq \varepsilon_n n^2/2. \tag{32}$$

Assume that there is a sequence of rows $i_1 < i_2 < \dots < i_p$ satisfying (32) with $i_m \geq i_{m-1} + 2$. By (23), we get

$$\left(w + \frac{1}{2}\right)n^2 \geq \frac{1}{2}\varepsilon_n n^2[(n - 1) + \dots + (n - 2p + 1)] = \frac{1}{2}\varepsilon_n n^2 p(n - p) \geq \varepsilon_n n^2 \frac{pn}{4}.$$

Thus, $p \leq \frac{4(w+\frac{1}{2})}{\varepsilon_n n}$. And so,

$$M \leq 2p \leq \frac{12w}{\varepsilon_n n}. \tag{33}$$

Since $|\mathcal{B}| \leq 3M \leq \frac{36w}{\varepsilon_n n}$, the number of ways to choose the set \mathcal{B} and the realizations of the rows with indices belonging to \mathcal{B} is at most

$$2^n (q^{n-1})^{36w/(\varepsilon_n n)} \leq 2^n q^{36w/\varepsilon_n} \leq n^{40w\varepsilon_n(\log q)/a^2} \leq n^{4 \cdot 10^3 w\varepsilon_n(\log q)/a}, \tag{34}$$

where $a = \varepsilon_n \sqrt{\log n}$ and in the last 2 inequalities, we used the assumption that $a \geq 10^{-2}$.

Step 2: Small rows. It’s left to control the set of small rows $[1, n - 1] \setminus \mathcal{B} =: \mathcal{S}$. We will use Lemma 3.6 for which we need to define the Ψ_i carefully.

Step 2.1: Definition of Ψ_i and r_i .

For $i \in \mathcal{B}$, we set $\Psi_i = \emptyset$. For $i \in \mathcal{S}$, we consider the cases that i is odd and even separately and define Ψ_i (and another quantity denoted by r_i , which is in essence, the skeleton size of row i) differently for each case.

Assume that $i \in \mathcal{S}$ is odd. Let I_i be the skeleton of row i and $r_i := |I_i|$. Note that by Lemma 3.3, the fact that $r_i \leq n/2$ and the assumption that row $i \in \mathcal{S}$, $r_i \leq \varepsilon_n n$. For an odd index i , let Ψ_i be the collection of nonzero intervals of row i with one or both endpoints belonging to the set $\mathcal{I}_i := I_i \cup (I_i \pm 1)$. By Lemma 3.3,

$$|\Psi_i| \geq r_i(n - r_i) \geq r_i(n - \varepsilon_n n). \tag{35}$$

Assume that $i \in \mathcal{S}$ is even. Note that the $\Psi_{i \pm 1}$ are defined above as $i \pm 1$ are odd. Let $\mathcal{K}_i = \mathcal{I}_{i-1}^c \cap \mathcal{I}_{i+1}^c$. Observe that any interval $[h, h']$ with $h, h' \in \mathcal{K}_i$, it holds that $[h, h']$ does not belong to $\Psi_{i-1} \cup \Psi_{i+1}$. Define Ψ_i to be the collection of nonzero intervals with both endpoints in \mathcal{K}_i . This guarantees that the assumption of Lemma 3.6 is satisfied.

To define r_i , let $\mathcal{L}_i = \mathcal{J}_{i-1}^c \cap \mathcal{J}_{i+1}^c$, where $\mathcal{J}_j := I_j \cup (I_j \pm 1) \cup (I_j \pm 2)$. We have $\mathcal{L}_i \subset \mathcal{K}_i$ and

$$|\mathcal{K}_i| \geq |\mathcal{L}_i| \geq n - 5|I_{i-1}| - 5|I_{i+1}| \geq n - 5r_{i-1} - 5r_{i+1} \geq n - 10\varepsilon_n n. \tag{36}$$

We consider a version of skeleton $\tilde{\mathcal{I}}_i = \{i_1, i_2, \dots, i_s\}$ restricted to the set \mathcal{L}_i as follows. Let $u = (y_{i,1}, \dots, y_{i,n-1}) \in (\mathbb{Z}/q\mathbb{Z})^{n-1}$ be the vector of row i . Let i_1 be the smallest index in \mathcal{L}_i such that $u_{i_1} \neq 0$. Let i_k be the smallest index such that $i_k \in \mathcal{L}_i$, $i_k \geq i_{k-1} + 2$ and $u_{i_k} \neq 0$ for all $2 \leq k \leq s$. Here s is the largest index for which this process has to stop,

meaning, $u_h = 0$ for all h satisfying both $h \geq i_s + 2$ and $h \in \mathcal{L}_i$. Set $r_i := s$. Observe that $r_i \leq \varepsilon_n n$ because $i \in \mathcal{S}$ and r_i is at most the size of the skeleton of row i .

By the same argument as in the proof of Lemma 3.3, the number of nonzero intervals with both endpoints belonging to the set $\{i_1, i_1 \pm 1, \dots, i_{r_i}, i_{r_i} \pm 1\}$ or one endpoint in this set and the other in \mathcal{K}_i is at least $r_i(|\mathcal{K}_i| - r_i)$. By the definition of \mathcal{L}_i , this set is a subset of \mathcal{K}_i , and so all such intervals belong to Ψ_i . Therefore,

$$|\Psi_i| \geq r_i(|\mathcal{K}_i| - r_i).$$

From that and (36), we get

$$|\Psi_i| \geq r_i(n - 10\varepsilon_n n - r_i) \geq r_i(n - 11\varepsilon_n n). \tag{37}$$

Applying Lemma 3.6 together with the equations (35) and (37), we have

$$(2w + 1)n^2 \geq 2N(y) \geq (|\Psi_1| + \dots + |\Psi_{n-1}|)n \geq \sum_{i \in \mathcal{S}} r_i(n - 11\varepsilon_n n),$$

which gives that

$$\sum_{i \in \mathcal{S}} r_i \leq \frac{2w + 1}{1 - 11\varepsilon_n} \leq 2w + 44w\varepsilon_n + 2. \tag{38}$$

Step 2.2: Realizations of small rows.

Choosing the r_i . The number of sequences $(r_i)_{i \in \mathcal{S}}$ of nonnegative integers satisfying (38) is at most the number of nonnegative integer solutions to $r_1 + r_2 + \dots + r_{n-1} = \mathbf{a}$ for some $\mathbf{a} \leq 2w + 44w\varepsilon_n + 2$ and is thus bounded by

$$\sum_{\mathbf{a}=0}^{2w+44w\varepsilon_n+2} \binom{\mathbf{a} + n - 2}{n - 2} \leq 7w \binom{6w + 2n}{n} \leq 7w \left(\frac{6ew + 2en}{n} \right)^n \leq n^{3 \cdot 10^4 \varepsilon_n w}, \tag{39}$$

where we used the assumptions that $\varepsilon_n \sqrt{\log n} \geq 10^{-2}$ and $w \geq \varepsilon_n n$.

Realizations of odd rows i . For each choice of the set \mathcal{B} , the realizations of rows j with $j \in \mathcal{B}$, and the sequence $(r_i)_{i \in \mathcal{S}}$, we need to bound the number of realizations of rows i with $i \in \mathcal{S}$.

If $i \in \mathcal{S}$ is odd, then by the definition of skeletons, the number of realizations of row i with a given r_i , the size of its skeleton, is at most

$$\binom{n}{r_i} q^{2r_i}. \tag{40}$$

Realizations of even rows i . Having chosen all of the rows i' with i' being odd, for each even index $i \in \mathcal{S}$, note that the set \mathcal{L}_i , defined earlier in the proof, is fixed as the rows $i \pm 1$ (odd) have been chosen. The number of choices for the restricted skeleton $\tilde{\mathcal{L}}_i$ is at most $\binom{n}{r_i}$. The number of realizations of row i inside the set \mathcal{L}_i with a given $\tilde{\mathcal{L}}_i$ is at most q^{2r_i} . The number of realizations of row i inside the set \mathcal{L}_i^c is at most $q^{|\mathcal{L}_i^c|} \leq q^{5r_{i-1} + 5r_{i+1}}$ by (36). Thus, the number of realizations of row i given $r_i, r_{i \pm 1}$ and the rows with odd indices is at most

$$\binom{n}{r_i} q^{2r_i + 5r_{i-1} + 5r_{i+1}} \tag{41}$$

Combining. From (40), and (41), the number of realizations of rows i with $i \in \mathcal{S}$ is at most

$$\begin{aligned} & \prod_{i \in \mathcal{S}, i=2k+1} \binom{n}{r_i} q^{2r_i} \prod_{i \in \mathcal{S}, i=2k} \binom{n}{r_i} q^{2r_i + 5r_{i-1} + 5r_{i+1}} \\ & \leq \exp\left(\log n \sum_{i \in \mathcal{S}} r_i + 12 \log q \sum_{i \in \mathcal{S}} r_i\right) \leq n^{2w + 3 \cdot 10^4 \varepsilon_n w + 500w\varepsilon_n (\log q)/a}, \end{aligned} \tag{42}$$

where in the last inequality, we used (38).

Combining (34), (39) and (42), we obtain the stated lemma. □

Finally, we are ready to prove (19) which completes the proof of the upper bound of Theorem 1.2.

Lemma 3.8. Let $\alpha_n = \log \log(16n)\sqrt{\log n}$. Let c be any number satisfying $c \geq \frac{640}{\log \log(16n)}$. Let $t = \frac{n^2}{4(1-\cos \frac{2\pi}{q})} \log n + \frac{n^2}{(1-\cos \frac{2\pi}{q})} 100c\alpha_n \log q$. We have for all $n \geq 4$ and $q \geq 2$,

$$\Sigma = \sum \left(1 - \frac{(1 - \cos 2\pi/q)N(y)}{\binom{n}{2}^2}\right)^{2t} \leq q^{-2c}/2, \tag{43}$$

where the sum is taken over all matrices $y \neq 0$ with $N(y) \leq \frac{\binom{n}{2}^2}{1-\cos \frac{2\pi}{q}}$.

Proof. Let $\varepsilon_n = \frac{1}{100\sqrt{\log n}}$ chosen with foresight. Let

$$w_1 = \max\{1, \varepsilon_n n\} \quad \text{and} \quad w_2 = \min\{w_1, 1/(100\varepsilon_n)\} = \min\{w_1, \sqrt{\log n}\} \geq 1.$$

We break up the left-hand side of (43) into sums T_1, T_2 where

- T_1 is the sum over all y with $1 \leq N(y) \leq (w_2 - 1/2)n^2$,
- T_2 is the sum over all y with $(w_2 - 1/2)n^2 < N(y) \leq \frac{\binom{n}{2}^2}{1-\cos 2\pi/q}$.

Observe that for these matrices y ,

$$0 \leq 1 - \frac{(1 - \cos 2\pi/q)N(y)}{\binom{n}{2}^2} \leq \exp\left(-\frac{4(1 - \cos 2\pi/q)N(y)}{n^4}\right).$$

Bound T_1 . Note that by combining Lemmas 3.3 and 3.4, every nonzero matrix y has $N(y) \geq (n - 1)^2 > (1 - 1/2)n^2$. Thus $T_1 \leq T_{1,1} + T_{1,2}$ where

$$T_{1,1} = \sum_{w=1}^{w_2-1} \left| \left\{ y : \left(w - \frac{1}{2}\right)n^2 < N(y) \leq (w - 40w\varepsilon_n)n^2 \right\} \right| \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - \frac{1}{2})}{n^2}\right),$$

$$T_{1,2} = \sum_{w=1}^{w_2-1} \left| \left\{ y : (w - 40w\varepsilon_n)n^2 < N(y) \leq \left(w + \frac{1}{2}\right)n^2 \right\} \right| \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - 40w\varepsilon_n)}{n^2}\right).$$

Note that if $w_2 = 1$ then $T_{1,1}$ and $T_{1,2}$ are just 0. Thus, for evaluating $T_{1,1}$ and $T_{1,2}$, it suffices to assume that $w_2 > 1$ which, by definition of w_2 , means $w_2 \leq \min\{1/(100\varepsilon_n), \varepsilon_n n\}$. This and $\varepsilon_n \leq \frac{1}{80}$ allow to apply the second part of Lemma 3.5 to get

$$T_{1,1} \leq \sum_{w=1}^{w_2-1} q^{16w} n^{2w-2} \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - \frac{1}{2})}{n^2}\right) \leq \sum_{w=1}^{\infty} q^{(16-4c\alpha)w} \leq \frac{q^{-2c}}{8}$$

and

$$T_{1,2} \leq \sum_{w=1}^{w_2-1} q^{32w} n^{2w} \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - 40w\varepsilon_n)}{n^2}\right) \leq \sum_{w=1}^{\infty} q^{(32-c\alpha/4)w} \leq \frac{q^{-2c}}{8},$$

where we use the fact that $\alpha \geq 1$ and $c\alpha \geq 640\sqrt{\log n}$. Thus, $T_1 \leq q^{-2c}/4$.

Bound T_2 . We have $T_2 \leq T_{2,1} + T_{2,2}$ where

$$T_{2,1} = \sum_{w=w_2}^{w_1-1} \left| \left\{ y : \left(w - \frac{1}{2}\right)n^2 < N(y) \leq \left(w + \frac{1}{2}\right)n^2 \right\} \right| \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - \frac{1}{2})}{n^2}\right),$$

$$T_{2,2} = \sum_{w=w_1}^{\infty} \left| \left\{ y : \left(w - \frac{1}{2}\right)n^2 < N(y) \leq \left(w + \frac{1}{2}\right)n^2 \right\} \right| \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - \frac{1}{2})}{n^2}\right).$$

As before, if $w_1 = 1$ then $T_{2,1}$ is simply 0. So to evaluate $T_{2,1}$, it suffices to assume that $w_1 = \varepsilon_n n > 1$, and so $w \leq w_1 - 1 \leq \varepsilon_n n$. This together with $\varepsilon_n \leq \frac{1}{80}$ allow to apply the first part of Lemma 3.5 to $T_{2,1}$ and get

$$T_{2,1} \leq \sum_{w=w_2}^{\infty} q^{32w} n^{2w+1+60\varepsilon_n w} \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - \frac{1}{2})}{n^2}\right) \leq \frac{q^{-2c}}{8},$$

where we used the fact that $c\alpha \geq 640\sqrt{\log n}$ and $w_2 \geq \frac{\sqrt{\log n}}{100}$.

Finally, since $w_1 \geq \varepsilon_n n$, applying Lemma 3.7 to $T_{2,2}$ yields

$$T_{2,2} \leq \sum_{w=w_1}^{\infty} n^{2w+C_1\varepsilon_n w} \exp\left(-\frac{8(1 - \cos 2\pi/q)t(w - \frac{1}{2})}{n^2}\right) \leq \frac{q^{-2c}}{8},$$

where C_1 is as in Lemma 3.7 and we used the fact that $C_1\varepsilon_n \leq \frac{6 \cdot 10^3 \log q}{\sqrt{\log n}}$, $w_1 \geq \frac{\sqrt{\log n}}{100}$ and $c\alpha \geq 640\sqrt{\log n}$. Thus, $T_2 \leq q^{-2c}/4$ which together with the bound on T_1 complete the proof. \square

Acknowledgement

We would like to thank Du Phan for helpful suggestions.

References

- [1] W. Baldoni-Silva, J. A. De Loera and M. Vergne. Counting integer flows in networks. *Found. Comput. Math.* **4** (2004) 277–314. MR2078665 <https://doi.org/10.1007/s10208-003-0088-8>
- [2] A. Barvinok. Brunn–Minkowski inequalities for contingency tables and integer flows. *Adv. Math.* **211** (2007) 105–122. MR2313530 <https://doi.org/10.1016/j.aim.2006.07.012>
- [3] A. Barvinok. Enumerating contingency tables via random permanents. *Combin. Probab. Comput.* **17** (2008) 1–19. MR2376421 <https://doi.org/10.1017/S0963548307008668>
- [4] A. Barvinok. Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes. *Int. Math. Res. Not. IMRN* (2009) 348–385.
- [5] A. Barvinok. What does a random contingency table look like? *Combin. Probab. Comput.* **19** (2010) 517–539. MR2647491 <https://doi.org/10.1017/S0963548310000039>
- [6] A. Barvinok and J. A. Hartigan. Maximum entropy Gaussian approximations for the number of integer points and volumes of polytopes. *Adv. in Appl. Math.* **45** (2010) 252–289. MR2646125 <https://doi.org/10.1016/j.aam.2010.01.004>
- [7] A. Barvinok and J. A. Hartigan. An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. *Trans. Amer. Math. Soc.* **364** (2012) 4323–4368. MR2912457 <https://doi.org/10.1090/S0002-9947-2012-05585-1>
- [8] A. Barvinok and J. A. Hartigan. The number of graphs and a random graph with a given degree sequence. *Random Structures Algorithms* **42** (2013) 301–348. MR3039682 <https://doi.org/10.1002/rsa.20409>
- [9] A. Barvinok, Z. Luria, A. Samorodnitsky and A. Yong. An approximation algorithm for counting contingency tables. *Random Structures Algorithms* **37** (2010) 25–66. MR2674620 <https://doi.org/10.1002/rsa.20301>
- [10] E. A. Bender. The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Math.* **10** (1974) 217–223. MR0389621 [https://doi.org/10.1016/0012-365X\(74\)90118-6](https://doi.org/10.1016/0012-365X(74)90118-6)
- [11] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A* **24** (1978) 296–307. MR0505796 [https://doi.org/10.1016/0097-3165\(78\)90059-6](https://doi.org/10.1016/0097-3165(78)90059-6)
- [12] I. Bezáková, N. Bhatnagar and E. Vigoda. Sampling binary contingency tables with a greedy start. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms* 414–423. ACM, New York, 2006. MR2368838 <https://doi.org/10.1145/1109557.1109604>
- [13] I. Bezáková, N. Bhatnagar and E. Vigoda. Sampling binary contingency tables with a greedy start. *Random Structures Algorithms* **30** (2007) 168–205. MR2283227 <https://doi.org/10.1002/rsa.20155>
- [14] I. Bezáková, A. Sinclair, D. Stefankovic and E. Vigoda. Negative examples for sequential importance sampling of binary contingency tables. In *Algorithms—ESA 2006* 136–147. *Lecture Notes in Comput. Sci.* **4168**. Springer, Berlin, 2006. MR2347138 https://doi.org/10.1007/11841036_15
- [15] I. Bezáková, A. Sinclair, D. Stefankovic and E. Vigoda. Negative examples for sequential importance sampling of binary contingency tables. *Algorithmica* **64** (2012) 606–620. MR2989466 <https://doi.org/10.1007/s00453-011-9569-3>
- [16] J. Blanchet and A. Stauffer. Characterizing optimal sampling of binary contingency tables via the configuration model. *Random Structures Algorithms* **42** (2013) 159–184. MR3019396 <https://doi.org/10.1002/rsa.20403>
- [17] J. Blitzstein and P. Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** (2010) 489–522. MR2809836 <https://doi.org/10.1080/15427951.2010.557277>
- [18] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** (1980) 311–316. MR0595929 [https://doi.org/10.1016/S0195-6698\(80\)80030-8](https://doi.org/10.1016/S0195-6698(80)80030-8)
- [19] E. R. Canfield, C. Greenhill and B. D. McKay. Asymptotic enumeration of dense 0–1 matrices with specified line sums. *J. Combin. Theory Ser. A* **115** (2008) 32–66. MR2378856 <https://doi.org/10.1016/j.jcta.2007.03.009>
- [20] Y. Chen, P. Diaconis, S. P. Holmes and J. S. Liu. Sequential Monte Carlo methods for statistical analysis of tables. *J. Amer. Statist. Assoc.* **100** (2005) 109–120. MR2156822 <https://doi.org/10.1198/016214504000001303>

- [21] Y. Chen, I. H. Dinwoodie and S. Sullivant. Sequential importance sampling for multiway tables. *Ann. Statist.* **34** (2006) 523–545. MR2275252 <https://doi.org/10.1214/009053605000000822>
- [22] F. R. K. Chung, R. L. Graham and S.-T. Yau. On sampling with Markov chains. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms* 55–77. Atlanta, GA, 1995, **9**, 1996. MR1611732 [https://doi.org/10.1002/\(SICI\)1098-2418\(199608/09\)9:1/2<155::AID-RSA4>3.3.CO;2-C](https://doi.org/10.1002/(SICI)1098-2418(199608/09)9:1/2<155::AID-RSA4>3.3.CO;2-C)
- [23] M. Cryan and M. Dyer. A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant. *J. Comput. System Sci.* **67** (2003) 291–310. Special issue on STOC2002 (Montreal, QC). MR2022833 [https://doi.org/10.1016/S0022-0000\(03\)00014-X](https://doi.org/10.1016/S0022-0000(03)00014-X)
- [24] 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. MR2231648 <https://doi.org/10.1137/S0097539703434243>
- [25] M. Cryan, M. Dyer and D. Randall. Approximately counting integral flows and cell-bounded contingency tables. *SIAM J. Comput.* **39** (2010) 2683–2703. MR2645886 <https://doi.org/10.1137/060650544>
- [26] P. Diaconis. *Group Representations in Probability and Statistics. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **11**. Institute of Mathematical Statistics, Hayward, CA, 1988. MR0964069
- [27] P. Diaconis. Threads Through Group Theory. In *Character Theory of Finite Groups. Contemp. Math.* **524**, 33–47. Amer. Math. Soc., Providence, RI, 2010. MR2731916 <https://doi.org/10.1090/conm/524/10343>
- [28] P. Diaconis and B. Efron. Testing for independence in a two-way table: New interpretations of the chi-square statistic. *Ann. Statist.* **13** (1985) 845–913. With discussions and with a reply by the authors. MR0803747 <https://doi.org/10.1214/aos/1176349634>
- [29] P. Diaconis and B. Efron. Probabilistic-geometric theorems arising from the analysis of contingency tables. In *Contributions to the Theory and Application of Statistics* 103–125. Academic Press, Boston, MA, 1987.
- [30] P. Diaconis and A. Gangolli. Rectangular arrays with fixed margins. In *Discrete Probability and Algorithms* 15–41. Minneapolis, MN, 1993. IMA Vol. Math. Appl. **72**. Springer, New York, 1995. MR1380519 https://doi.org/10.1007/978-1-4612-0801-3_3
- [31] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.* **21** (1993) 2131–2156. MR1245303
- [32] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** (1993) 696–730. MR1233621
- [33] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** (1981) 159–179. MR0626813 <https://doi.org/10.1007/BF00535487>
- [34] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** (1998) 363–397. MR1608156 <https://doi.org/10.1214/aos/1030563990>
- [35] M. Dyer and C. Greenhill. Polynomial-time counting and sampling of two-rowed contingency tables. *Theoret. Comput. Sci.* **246** (2000) 265–278. MR1780242 [https://doi.org/10.1016/S0304-3975\(99\)00136-X](https://doi.org/10.1016/S0304-3975(99)00136-X)
- [36] M. Dyer, R. Kannan and J. Mount. Sampling contingency tables. *Random Structures Algorithms* **10** (1997) 487–506. MR1608222 [https://doi.org/10.1002/\(SICI\)1098-2418\(199707\)10:4<487::AID-RSA4>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1098-2418(199707)10:4<487::AID-RSA4>3.0.CO;2-Q)
- [37] C. Greenhill and B. D. McKay. Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums. *Adv. in Appl. Math.* **41** (2008) 459–481. MR2459445 <https://doi.org/10.1016/j.aam.2008.01.002>
- [38] D. Hernek. Random generation of $2 \times n$ contingency tables. *Random Structures Algorithms* **13** (1998) 71–79. MR1634343 [https://doi.org/10.1002/\(SICI\)1098-2418\(199808\)13:1<71::AID-RSA4>3.3.CO;2-O](https://doi.org/10.1002/(SICI)1098-2418(199808)13:1<71::AID-RSA4>3.3.CO;2-O)
- [39] A. Liebenau and N. Wormald. Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph, 2017. Preprint. Available at [arXiv:1702.08373](https://arxiv.org/abs/1702.08373).
- [40] T. Matsui, Y. Matsui and Y. Ono. Random generation of $2 \times 2 \times \dots \times 2 \times J$ contingency tables. *Theoret. Comput. Sci.* **326** (2004) 117–135. MR2094245 <https://doi.org/10.1016/j.tcs.2004.06.014>
- [41] B. D. McKay and N. C. Wormald. Asymptotic enumeration by degree sequence of graphs of high degree. *European J. Combin.* **11** (1990) 565–580. MR1078713 [https://doi.org/10.1016/S0195-6698\(13\)80042-X](https://doi.org/10.1016/S0195-6698(13)80042-X)
- [42] B. D. McKay and N. C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica* **11** (1991) 369–382. MR1137769 <https://doi.org/10.1007/BF01275671>
- [43] B. D. McKay and N. C. Wormald. The degree sequence of a random graph. I. The models. *Random Structures Algorithms* **11** (1997) 97–117. MR1610253 [https://doi.org/10.1002/\(SICI\)1098-2418\(199709\)11:2<97::AID-RSA1>3.3.CO;2-E](https://doi.org/10.1002/(SICI)1098-2418(199709)11:2<97::AID-RSA1>3.3.CO;2-E)
- [44] B. Morris. Improved bounds for sampling contingency tables. In *Randomization, Approximation, and Combinatorial Optimization* 121–129. Berkeley, CA, 1999. *Lecture Notes in Comput. Sci.* **1671**. Springer, Berlin, 1999. MR1775514 https://doi.org/10.1007/978-3-540-48413-4_12
- [45] G. W. Soules. New permanent upper bounds for nonnegative matrices. *Linear Multilinear Algebra* **51** (2003) 319–337. MR2010711 <https://doi.org/10.1080/0308108031000098450>
- [46] D. B. Wilson. Mixing times of Lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.* **14** (2004) 274–325. MR2023023 <https://doi.org/10.1214/aoap/1075828054>
- [47] N. C. Wormald. *Models of Random Regular Graphs. London Mathematical Society Lecture Note Series*, 1999. MR1725006