PHASE TRANSITION IN RANDOM CONTINGENCY TABLES
WITH NON-UNIFORM MARGINS

SAM DITTMER, HANBAEK LYU, AND IGOR PAK

ABSTRACT. For parameters $n, \delta, B,$ and C, let $X = (X_{k\ell})$ be the random uniform contingency table whose first $\lfloor n^\delta \rfloor$ rows and columns have margin $\lfloor BCn \rfloor$ and the last n rows and columns have margin $\lfloor Cn \rfloor$. For every $0 < \delta < 1$, we establish a sharp phase transition of the limiting distribution of each entry of X at the critical value $B_c = 1 + \sqrt{1 + 1/C}$. In particular, for $1/2 < \delta < 1$, we show that the distribution of each entry converges to a geometric distribution in total variation distance, whose mean depends sensitively on whether $B < B_c$ or $B > B_c$. Our main result shows that $\mathbb{E}[X_{11}]$ is uniformly bounded for $B < B_c$, but has sharp asymptotic $C(B - B_c) n^{1-\delta}$ for $B > B_c$. We also establish a strong law of large numbers for the row sums in top right and top left blocks.

1. INTRODUCTION

1.1. **Random contingency tables.** Contingency tables are fundamental objects in statistics for studying dependence structure between two or more variables, see e.g. [Eve92, FLL17, Kat14]. They also correspond to bipartite multi-graphs with given degrees and play an important role in combinatorics and graph theory, see e.g. [Bar09, DG95, DS98]. Random contingency tables have been intensely studied in a variety of regimes, yet remain largely out of reach in many interesting special cases, see e.g. [Bar10b, CM10].

Let $\mathbf{r} = (r_1, \ldots, r_m) \in \mathbb{N}^m, \mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{N}^n$ be two nonnegative integer vectors with the same sum of entries. Denote by $\mathcal{M}(\mathbf{r}, \mathbf{c})$ the set of all $(n \times m)$ contingency tables with row sums r_i and column sums c_j, i.e.

$$(1.1) \qquad \mathcal{M}(\mathbf{r}, \mathbf{c}) := \left\{ (a_{ij}) \in \mathbb{N}^{mn} \ \middle| \ \sum_{k=1}^{n} a_{ik} = r_i, \ \sum_{k=1}^{m} a_{kj} = c_j \ \text{ for all } 1 \le i \le n, 1 \le j \le m \right\}.$$

Let $X = (X_{ij})$ be the contingency table chosen uniformly at random from $\mathcal{M}(\mathbf{r}, \mathbf{c})$. The asymptotic properties of the entries of X as $m, n \to \infty$ is the subject of this paper. When the margins are uniform, i.e. $r_1 = \ldots = r_m$ and $c_1 = \ldots = c_n$, the exact asymptotics for $|\mathcal{M}(\mathbf{r}, \mathbf{c})|$ are known [CM10, GM08]. In fact, the distribution of individual entries X_{ij} is asymptotically geometric and the dependence between the entries vanish as the size of the table goes to infinity [CDS10].

In this paper we analyze random square contingency table with the first $\lfloor n^\delta \rfloor$ row and column margins $\lfloor BCn \rfloor$, and the last n row and column margins $\lfloor Cn \rfloor$. Viewing such X as block matrices, see Figure 1, it is natural to assume that the entries are again nearly independent and identically distributed within each block. However, there is still one degree of freedom remaining: the distribution of mass of each block. We establish a sharp phase transition for this distribution. The following corollary is a special case of general results we present in the next section.

Corollary 1.1 (see Theorem 2.2)**.** *Fix constants $B, C > 0$ and $1/2 < \delta < 1$. Let $X = (X_{ij})$ be the uniform random contingency table with the first $\lfloor n^\delta \rfloor$ row and column margins $\lfloor BCn \rfloor$, and the last n row and column margins $\lfloor Cn \rfloor$. Then:*

$$\mathbb{E}[X_{11}] = \Theta(1) \ \text{ for } B < B_c, \quad \text{and} \quad \mathbb{E}[X_{11}] = \Theta(n^{1-\delta}) \ \text{ for } B > B_c,$$

where the critical value $B_c = 1 + \sqrt{1 + 1/C}$.

In the next section we present various extensions and refinements of this result, including determining the precise constants implied by the Θ notation. We also extend the results to $0 < \delta \le 1/2$, although our results are not as strong in this case.
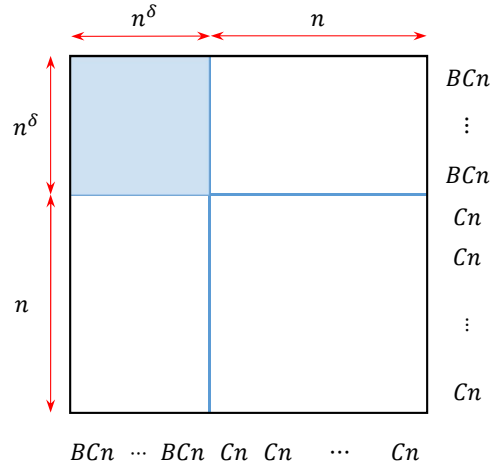
FIGURE 1. Contingency table with parameters n, δ, B and C. First $\lfloor n^\delta \rfloor$ rows and columns have margins $\lfloor Cn \rfloor$, the last n rows and columns have margins $\lfloor BCn \rfloor$.

The story behind the phase transition in the corollary is quite interesting. For $\delta = 0$, $B = 3$ and $C = 1$ the phenomenon of large X_{11} entry was first observed by Barvinok in [Bar10b, §1.5] for the (non-uniform) distribution of *"typical" contingency tables*. In fact, Barvinok showed that for $\delta = 0$ and $C = 1$, the phase transition for typical tables happens at $B_c = 1 + \sqrt{2}$, see [Bar10b, Bar12].

In [DLP19+], the authors tested empirically uniform contingency tables using a new MCMC algorithm introduced in [DP18], and the experiments seem to confirm Barvinok's conjectured value for the critical B_c. In fact, the simulations show drastically different behavior for the subcritical $B < B_c$ vs. supercritical $B > B_c$ cases. Here we analyze the X_{11} entry in a random uniform $n \times n$ contingency table with both margins $(Bn, n, \ldots, n)$, see Figure 2. This is the case $C = 1$, where $B_c = 1 + \sqrt{2} \approx 2.41$. In some sense, the simulations
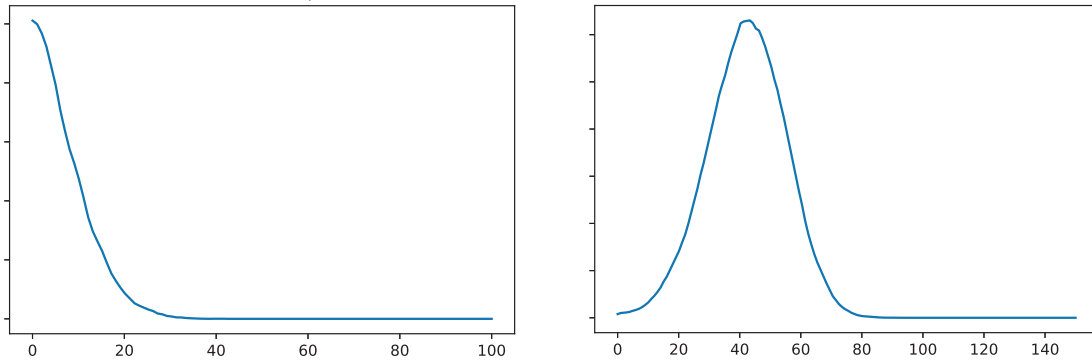

FIGURE 2. Summary of 10^4 simulations of X_{11} entry in a random uniform $n \times n$ contingency table $X = (X_{ij})$ with both margins $(Bn, n, \ldots, n)$, where $n = 50$. The left graph corresponds to a subcritical value $B = 2$, and the right graph to a supercritical value $B = 3$.

showed a sharper phase transition than the typical matrices: not only the expectation $\mathbb{E}[X_{11}]$ exhibited jump from bounded to having linear growth, but the distribution of X_{11} switches from geometric in the subcritical case to normal in the supercritical case.

This paper gives the first rigorous proof of the phase transition in the uniform case. Although we do not cover the $\delta = 0$ case introduced by Barvinok, we conjecture the phase transition extends to this case. In fact, our results go beyond what the simulations in [DLP19+] suggest, as we interpolate between $\delta = 0$ and (uniform) $\delta = 1$ case. Rather surprisingly, we show that for $\delta > 1/2$ the behavior of random uniform and typical matrices remains similar, with a geometric distribution in the supercritical case. We conjecture that there is an additional phase transition at $\delta = 1/2$, and for $\delta < 1/2$ the distribution of X_{11} is normal in the supercritical case (see Conjecture 3.2). In the limiting case $\delta = 0$, this is supported by the simulations mentioned above. Further conjectures with more refined estimates are given in Section 3.

1.2. **Background.** Let $\mathscr{P}(\mathbf{r}, \mathbf{c}) \subseteq \mathbb{R}_+^{mn}$ be the *transportation polytope* of *real* nonnegative contingency tables with margins $\mathbf{r}$ and $\mathbf{c}$, i.e. defined by (1.1) over $\mathbb{R}_+$. Clearly, $\mathscr{M}(\mathbf{r}, \mathbf{c}) = \mathscr{P}(\mathbf{r}, \mathbf{c}) \cap \mathbb{Z}^{mn}$. When $r_1 = \ldots = r_n = c_1 = \ldots = 1$, $m = n$, we have $\mathscr{P}(\mathbf{r}, \mathbf{c})$ is the classical *Birkhoff polytope*, of interest in Combinatorics, Discrete Geometry, Combinatorial Optimization and Discrete Probability, see e.g. [DK14, Pak00]. The asymptotic behavior of the $\mathrm{vol}(\mathscr{P}(\mathbf{r}, \mathbf{c}))$ is known [CM09], as well as the exact value and the whole Ehrhart polynomial for $n \leq 10$, see [BP03], and numerical estimates for $11 \leq n \leq 14$ [CV16]. Such sharp volume estimates were crucially used in [CDS10] to analyze the asymptotic behavior of random contingency table X for uniform margins.

For non-uniform margins the existing sharp asymptotic results cover only *smooth margins*, a technical condition which includes the case when all ratios r_i/r_k and c_j/c_ℓ are bounded, see [BLSY10, BBK72, BC78, CM10] for precise statements. For general margins $\mathbf{r}$ and $\mathbf{c}$, the upper and lower bounds on $|\mathscr{M}(\mathbf{r}, \mathbf{c})|$ were given in [Bar09, Bar10b] (see also Theorem 4.3). The proofs of our main results rely heavily on these bounds.

Now, in statistics, a popular practice is to sample Y from the *hypergeometric* (Fisher–Yates) *distribution* defined as follows:

$$(1.2) \qquad \mathbb{P}(Y = (y_{ij})) := \phi(\mathbf{r}, \mathbf{c}) \cdot \prod_{i=1}^{m} \prod_{j=1}^{n} \frac{1}{y_{ij}!}, \quad \text{where} \quad \phi(\mathbf{r}, \mathbf{c}) = \frac{1}{N!} \prod_{i=1}^{m} r_i! \prod_{j=1}^{n} c_j!$$

and $N = r_1 + \ldots + r_m = c_1 + \ldots + c_n$ denotes the total sum of a contingency table in $\mathscr{M}(\mathbf{r}, \mathbf{c})$. We refer to [DE85, Eve92] for an extensive discussion and to [FLL17, Kat14] for the recent treatment.

The rationale behind this approach lies in the *independence table* $W = (w_{ij}) \in \mathscr{P}(\mathbf{r}, \mathbf{c})$, defined by $w_{ij} := (r_i c_j)/N$. This table W gives both the expectation of the Fisher–Yates distribution and is also the unique maximizer of the following strictly concave function

$$(1.3) \qquad \mathrm{H}(W) = \sum_{i,j} \frac{w_{ij}}{N} \log \frac{N}{w_{ij}}$$

in the transportation polytope $\mathscr{P}(\mathbf{r}, \mathbf{c})$, see [Goo63, Ex. (iv)]. Note that for each $Q \in \mathscr{P}(\mathbf{r}, \mathbf{c})$, if we view Q/N as the 'population contingency table', where each entry q_{ij}/N is understood as the marginal probability of the entry (i, j), then $\mathrm{H}(W)$ defined at (1.3) is the entropy of the probability mass function Q/N. However, it is known that the hypergeometric distribution may not properly capture the behavior of the uniform contingency table $X \in \mathscr{M}(\mathbf{r}, \mathbf{c})$.

When one tries to find the marginal distribution for Y_{ij} that maximizes the overall entropy subject to the margin condition, instead of viewing each contingency table as a rescaled probability mass function, one finds that the entries must be independent and geometrically distributed. Furthermore, one can further maximize the entropy by optimizing the mean of each entry. This leads to the notion of the *typical table* $Z = (z_{ij}) \in \mathscr{P}(\mathbf{r}, \mathbf{c})$ (see Definition 4.1), introduced by Barvinok in [Bar09] and further exploited in [Bar10a, Bar10b, BH10]. The behavior of typical and independence tables are known to be similar when the margins are relatively uniform [BLSY10], but could be drastically different when the margin are strongly asymmetric [Bar10b, §1.6].

In [Bar10b], Barvinok showed that there exists a phase transition in the behavior of the typical table for a simple model of contingency tables with asymmetric margins. Namely, let $Z = (z_{ij})$ be the typical table for $\mathcal{M}(\mathbf{r}, \mathbf{c})$ where $\mathbf{r} = \mathbf{c} := (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \ldots, \lfloor Cn \rfloor) \in \mathbb{N}^{n+1}$. For $B = 1$, all entries of Z are equal by the symmetry. In particular, the corner entry z_{11} is bounded by C for all n. On the other hand, Barvinok [Bar10b] showed that for $B > 1 + \sqrt{2}$, the entry z_{11} has linear growth

$$z_{11} \geq (B - 1 - \sqrt{2})\, n \quad \text{for all} \ \ n \geq 1,$$

while all the other entries of Z are uniformly bounded by $1 + \sqrt{2}$. Hence, as B passes a certain critical value $\leq 1 + \sqrt{2}$, the "mass" within the typical table Z suddenly concentrates at the corner entry z_{11}. As we mentioned earlier, this was the starting observation of the paper.

1.3. **Notation.** We use $\mathbb{N} = \{0, 1, 2, \ldots\}$ and $\mathbb{R}_+ = \{x \in \mathbb{R},\, x \geq 0\}$. For all $a, b \in \mathbb{R}$, denote $a \wedge b := \min(a, b)$ and $a^+ := \max(a, 0)$.

For all $\lambda > 0$, we write $Y \sim \mathrm{Geom}(\lambda)$ for a discrete random variable Y with probability mass function[1]

$$\mathbb{P}(Y = k) := \left(\frac{1}{1 + \lambda} \right) \left(\frac{\lambda}{1 + \lambda} \right)^k, \quad \text{for all} \ \ k \in \mathbb{N}.$$

Note that $\mathbb{E}[Y] = \lambda$. We call Y a *geometric random variable* with mean λ. For every two probability distributions μ_1, μ_2 over a countable sample space Ω, the *total variation distance* is defined as

$$d_{TV}(\mu_1, \mu_2) := \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)|.$$

Let X and Y be random variables with distribution μ_1 and μ_2, respectively. To simplify the notation, we write:

$$d_{TV}(\mu_1, \mu_2) = d_{TV}(X, \mu_2) = d_{TV}(\mu_1, Y) = d_{TV}(X, Y).$$

2. Statement of results

For parameters $n \geq 1$, $0 \leq \delta \leq 1$, and $B, C \geq 0$, let $\mathcal{M}_{n,\delta}(B, C) = \mathcal{M}(\mathbf{r}, \mathbf{c})$, where

$$(2.1) \qquad\qquad \mathbf{r} = \mathbf{c} := \left(\lfloor BCn \rfloor, \ldots, \lfloor BCn \rfloor, \lfloor Cn \rfloor, \ldots, \lfloor Cn \rfloor \right) \in \mathbb{N}^{\lfloor n^\delta \rfloor + n}.$$

In other words, $\mathcal{M}_{n,\delta}(B, C)$ is the set of contingency tables whose first $\lfloor n^\delta \rfloor$ rows and columns have margin $\lfloor BCn \rfloor$ and the other n rows and columns have margin $\lfloor Cn \rfloor$, see Figure 1. Let $X = (X_{ij})$ be the random contingency table sampled uniformly from $\mathcal{M}_{n,\delta}(B, C)$. We are interested in the asymptotic behavior of the entry X_{ij} as $n \to \infty$ for various choice of parameters δ, B and C. Note that the entries within each of the four blocks in Figure 1 have the same distribution by the symmetry. Hence it is sufficient to restrict the attention to the entries X_{11}, $X_{1,n+1}$, and $X_{n+1,n+1}$.

We establish a sharp phase transition at

$$B_c := 1 + \sqrt{1 + 1/C}$$

for the limiting expectation of the entries of X. The following theorem shows that the limiting distribution of each entry of X is geometric with mean depending on whether $B < B_c$ or $B > B_c$; see Figure 3.

Theorem 2.1. *Fix constants $\delta, B, C > 0$, and let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B, C)$ uniformly at random. Fix $\varepsilon > 0$ and let B_c be as above.*

(i) [bottom right] *For all $0 \leq \delta \leq 1$ and $B, C > 0$, we have:*

$$d_{TV}(X_{n+1,n+1}, \mathrm{Geom}(C)) = O(n^{\delta - 1} + n^{-1/2 + \varepsilon}).$$

---

[1]This notation is somewhat nonstandard, but is more convenient for our purposes.

(ii) [sides] *For all $0 < \delta \le 1$ and $B \ne B_c$, we have:*
$$d_{TV}\big(X_{1,n+1}, \mathrm{Geom}(R)\big) = O\big(n^{\delta-1} + n^{-\delta/2+\varepsilon}\big),$$

where
$$R = \begin{cases} BC & \text{if } B < B_c, \\ B_c C & \text{if } B > B_c. \end{cases}$$

(iii) [top left] *For all $1/2 < \delta < 1$ and $B \ne B_c$, we have:*
$$d_{TV}\big(X_{11}, \mathrm{Geom}(R)\big) = O\big(n^{\delta-1} + n^{1/2-\delta+\varepsilon}\big),$$

where
$$R = \begin{cases} \frac{B^2(1+C)}{(B_c-B)(B+B_c-2)} & \text{if } B < B_c, \\ C(B-B_c)n^{1-\delta} + O(1) & \text{if } B > B_c. \end{cases}$$
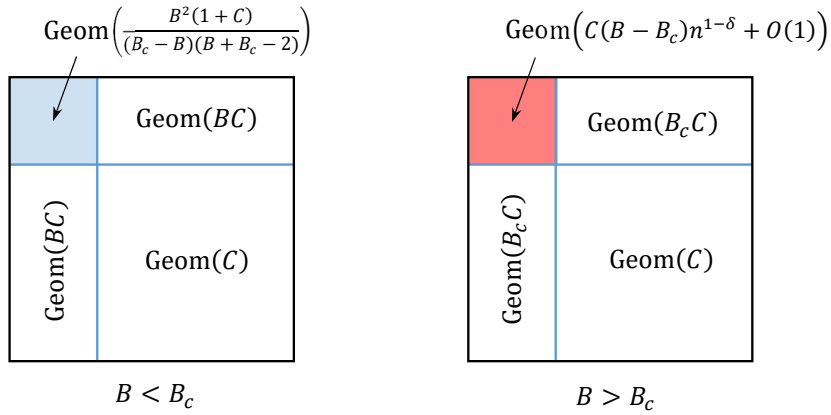

FIGURE 3. Limiting distributions of the entries in the uniform contingency table $X \in \mathcal{M}_{n,\delta}(B,C)$ in the subcritical $B < B_c$ (left) and supercritical $B > B_c$ (right) regimes for thick bezels $1/2 < \delta < 1$.

Our second result proves the phase transition of entries of random contingency tables $X \in \mathcal{M}_{n,\delta}(B,C)$ in expectation.

Theorem 2.2. *Fix constants $\delta, B, C > 0$, and let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B,C)$ uniformly at random. Let B_c be as above.*

(i) [bottom right] *For all $0 \le \delta \le 1$, $B, C > 0$, and $\varepsilon > 0$, we have:*
$$|\mathbb{E}[X_{n+1,n+1}] - C| = O(n^{\delta-1+\varepsilon}).$$

(ii) [sides] *For all $\alpha, \varepsilon > 0$, we have:*
$$\begin{cases} |\mathbb{E}[X_{1,n+1}] - BC| = O\big(n^{\delta-1} + n^{-(\delta/2)+\varepsilon}\big) & \text{if } B < B_c \text{ and } 0 < \delta < 1, \\ |\mathbb{E}[X_{1,n+1} \wedge n^\alpha] - B_c C| = O\big(n^{\delta-1} + n^{\alpha-(\delta/2)+\varepsilon}\big) & \text{if } B > B_c \text{ and } 0 < \delta < 1, \\ |\mathbb{E}[X_{1,n+1}] - B_c C| = O\big(n^{(1/2)-\delta+\varepsilon}\big) & \text{if } B > B_c \text{ and } 1/2 < \delta < 1. \end{cases}$$

(iii) [top left] *For all $\alpha, \varepsilon > 0$, we have:*
$$\begin{cases} |n^{\delta-1}\mathbb{E}[X_{11}]| = O\big(n^{\delta-1} + n^{-\delta+\varepsilon}\big) & \text{if } B < B_c \text{ and } 0 < \delta < 1, \\ \left|\mathbb{E}[X_{11} \wedge n^\alpha] - \frac{B^2(1+C)}{(B_c-B)(B+B_c-2)}\right| = O\big(n^{\alpha+(1/2)-\delta+\varepsilon}\big) & \text{if } B < B_c \text{ and } 1/2 < \delta < 1, \end{cases}$$

and

$$\begin{cases} \big||C(B - B_c) - n^{\delta-1}\mathbb{E}[X_{11}] - \mathbb{E}\big[(X_{1,n+1} - n^{\alpha})^+\big]\big| = O\big(n^{\delta-1} + n^{\alpha-(\delta/2)+\varepsilon}\big) & \text{if } B > B_c \text{ and } 0 < \delta < 1, \\ \big||C(B - B_c) - n^{\delta-1}\mathbb{E}[X_{11}]\big| = O\big(n^{(1/2)-\delta+\varepsilon}\big) & \text{if } B > B_c \text{ and } 1/2 < \delta < 1. \end{cases}$$

Lastly, we establish strong law of large numbers for row sums of entries in X in the top right and bottom right blocks.

Theorem 2.3. *Fix $B, C > 0$ and $1/2 < \delta < 1$. Let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B, C)$ uniformly at random, and let B_c be as above. Then a.s., as $n \to \infty$, we have:*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_{1,k+\lfloor n^{\delta}\rfloor} = \begin{cases} BC & \text{if } B < B_c, \\ B_c C & \text{if } B > B_c. \end{cases}$$

Furthermore, for all $B, C > 0$ and $0 \le \delta < 1$, a.s. as $n \to \infty$, we have:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_{n+1,k+\lfloor n^{\delta}\rfloor} = C.$$

As we mentioned in the introduction, the proofs utilize Barvinok's technology of "typical tables" and upper and lower bounds on the number $|\mathcal{M}(\mathbf{r}, \mathbf{c})|$ of contingency tables for general margins. Except for the next section where we summarize conjectural extensions of our theorems, the rest of the paper is dedicated to proofs of the theorems.

3. CONJECTURES

In view of the strong law of large numbers for the top right block of X given by Theorem 2.3, we conjecture that a central limit theorem also holds for the top right block in the supercritical regime $B > B_c$ for at least when $1/2 < \delta < 1$. However, we believe that a central limit behavior should not be expected for the subcritical regime $B < B_c$. Our rationale is that, according to Theorem 2.3, the first row sum in the top right block of X is asymptotically BCn for $B < B_c$, which is the full row sum of X. Hence there is not much room for each entry in the top right block to fluctuate. On the other hand, for $B > B_c$, the row sum in the top right block contributes only to $B_c Cn$, which is independent of B. Hence when $B \gg B_c$ is large, there is enough room for them to fluctuate, and they would not feel the 'bar' of BC since they must fluctuate around $B_c C \ll BC$.

Conjecture 3.1. *Fix $B, C > 0$ and $0 < \delta < 1$. Let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B, C)$ uniformly at random. Denote*

$$S_{n,\delta}(B, C) := \sum_{k=1}^{n} X_{1,k+\lfloor n^{\delta}\rfloor}.$$

Then as $n \to \infty$, we have:

$$\begin{cases} \frac{1}{\sqrt{n}}\Big(S_{n,\delta}(B, C) - BCn\Big) \longrightarrow 0 & \text{a.s.} & \text{if } B < B_c, \\ \frac{1}{\sqrt{n}}\Big(S_{n,\delta}(B, C) - B_c Cn\Big) \Longrightarrow \sqrt{B_c C + (B_c C)^2} \cdot \mathcal{N}(0, 1) & \text{if } B > B_c, \end{cases}$$

where $\mathcal{N}(0, 1)$ is the standard normal distribution and "$\Rightarrow$" denotes the weak convergence.

Note that $B_c C + (B_c C)^2$ is the variance of the geometric distribution with mean $B_c C$. We remark that currently we only know that, for all $\alpha < \delta/3$ and $B > B_c$, we have:

$$\mathbb{E}\big[X_{1,n+1}^2 \wedge n^{\alpha}\big] \longrightarrow (B_c C)^2 \quad \text{as } n \to \infty.$$

Unfortunately, we are not able to get rid of the truncation in this formula, in contrast to Lemma 3.4 in [CDS10] for the uniform margin case. This is partly because our argument relies on the loose estimates of

$|\mathcal{M}(\mathbf{r}, \mathbf{c})|$ given by Theorem 4.3. Replacing the LHS with $\mathbb{E}[X_{1,n+1}^2]$ would be the first step in proving Conjecture 3.1.

Next, we conjecture that there exists a phase transition in δ with respect to the limiting distribution of X_{11} in the supercritical regime $B > B_c$. For the *thick bezel case* $1/2 < \delta < 1$, Theorem 2.1 shows that $n^{\delta-1} X_{11}$ converges in distribution to a geometric random variable. For the *thin bezel case* $0 < \delta < 1/2$, we conjecture that it should converge to a normal distribution. Roughly speaking, the sum of n^δ terms $X_{11} + \ldots + X_{1,\lfloor n^\delta \rfloor}$ is asymptotically a normal random variable by Conjecture 3.1. Thus, if $\delta < 1/2$, then there are not enough terms in this sum to exhibit central limit behavior. Hence the limiting distribution of this sum should be some rescaled version of the marginal distribution of X_{11}.

To make a more precise conjecture, let $S_{n,\delta}(B, C)$ be as in Conjecture 3.2. Then we write:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n^\delta \rfloor} \left[X_{1,k} - C(B - B_c) n^{1-\delta} \right] = \frac{1}{\sqrt{n}} \left(S_{n,\delta}(B, C) - B_c C n \right).$$

Assuming the summands in the left hand side are asymptotically uncorrelated, taking variance in each side gives

$$\mathrm{Var}(X_{11}) \sim n^{1-\delta} \left(B_c C + (B_c C)^2 \right).$$

Hence we have at the following conjecture.

Conjecture 3.2. *Fix $B, C > 0$ and $0 < \delta < 1/2$. Let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B, C)$ uniformly at random. Then*

$$\frac{X_{11} - C(B - B_c) n^{1-\delta}}{n^{(1-\delta)/2} \sqrt{B_c C + (B_c C)^2}} \implies \mathcal{N}(0, 1).$$

Further conjectures and open problems are given in Section 9.

4. CONCENTRATION IN BLOCKS

As we mentioned in the introduction, Barvinok [Bar10b] introduced the notion of *typical table* for general contingency tables, which captures some "typical behavior" of the uniform random contingency table of fixed margins. We start with the precise definition:

Definition 4.1 (Typical table). Fix margins $\mathbf{r} \in \mathbb{N}^m$ and $\mathbf{c} \in \mathbb{N}^n$. Let $\mathscr{P}(\mathbf{r}, \mathbf{c}) \subseteq \mathbb{R}_+^{mn}$ denote the transportation polytope. For each $X = (X_{ij}) \in \mathscr{P}(\mathbf{r}, \mathbf{c})$, define

(4.1) $$g(X) = \sum_{1 \le i,j \le n} f(X_{ij}),$$

where the function $f : [0, \infty) \to [0, \infty)$ is defined by

$$f(x) = (x + 1) \log(x + 1) - x \log x.$$

The *typical table* $Z \in \mathscr{P}(\mathbf{r}, \mathbf{c})$ for $\mathcal{M}(\mathbf{r}, \mathbf{c})$ is defined by

$$Z = \underset{X \in \mathscr{P}(\mathbf{r}, \mathbf{c})}{\arg\max}\, g(X).$$

Here $\mathscr{P}(\mathbf{r}, \mathbf{c})$ is is the transportation polytope of margins $\mathbf{r}$ and $\mathbf{c}$ defined in the introduction. Since the function g defined at (4.1) is strictly concave, it attains a unique maximizer on the transportation polytope $\mathscr{P}(\mathbf{r}, \mathbf{c})$ and thus the typical table is well-defined.

One of the building blocks of our main result is Lemma 4.2, which says that the law of an entry in a large block of random contingency table is attracted toward a geometric distribution, whose mean is dictated by

the corresponding typical table. Given the set $\mathcal{M}(\mathbf{r},\mathbf{c})$ of $m \times n$ contingency tables of margins $\mathbf{r}$ and $\mathbf{c}$, we call a set $\mathcal{B} \subseteq \{1,\dots,m\} \times \{1,\dots,n\}$ of indices a *block* of $\mathcal{M}(\mathbf{r},\mathbf{c})$ if

$$(r_i, c_j) = (r_{i'}, c_{j'}) \quad \text{for all } (i,j), (i',j') \in \mathcal{B}.$$

Observe that when $X = (X_{ij})$ is sampled from $\mathcal{M}(\mathbf{r},\mathbf{c})$ uniformly at random and $\mathcal{B}$ is a block of $\mathcal{M}(\mathbf{r},\mathbf{c})$, the entries X_{ij} for all $(i,j) \in \mathcal{B}$ have the same distribution by the symmetry. Moreover, the entries of the typical table Z within a block are the same.

Lemma 4.2. *Let $\mathcal{M}(\mathbf{r},\mathbf{c})$ be the set of all $m \times n$ contingency tables of margins $\mathbf{r}$ and $\mathbf{c}$. Let $X = (X_{ij})$ be sampled from $\mathcal{M}(\mathbf{r},\mathbf{c})$ uniformly at random. Suppose $\mathcal{B}_1,\dots,\mathcal{B}_k$ are (not necessarily distinct) blocks in $\mathcal{M}(\mathbf{r},\mathbf{c})$ with $|\mathcal{B}_1| \le \dots \le |\mathcal{B}_k|$. Then there exists an absolute constant $\gamma > 0$, s.t. for each $I = (I_1,\dots,I_k) \in \mathcal{B}_1 \times \dots \times \mathcal{B}_k$ and $t > 0$, we have:*

$$d_{TV}\left(\prod_{\ell=1}^{k} X_{I_\ell}, \prod_{\ell=1}^{k} Y_{I_\ell} \right) \le t + N^{\gamma(m+n)} \exp\left(-\left\lfloor \frac{|\mathcal{B}_1|}{k} \right\rfloor \frac{t^2}{2} \right),$$

where $Z = (z_{ij})$ is the typical table for $\mathcal{M}(\mathbf{r},\mathbf{c})$, and $Y = (Y_J)$ is the random matrix of independent entries with $Y_{ij} \sim \mathrm{Geom}(z_{ij})$, and $N = r_1 + \dots + r_m = c_1 + \dots + c_n$.

Our proof of Lemma 4.2 relies upon the following results of Barvinok, see [Bar10b, Thm 1.7], [Bar09, Thm 1.1], and [Bar09, Lem 1.4].

Theorem 4.3 ([Bar09, Bar10b]). *Fix margins $\mathbf{r} \in \mathbb{N}^m$ and $\mathbf{c} \in \mathbb{N}^n$. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}(\mathbf{r},\mathbf{c})$, and let $Y = (Y_{ij})$ be the $(m \times n)$ random matrix of independent entries where $Y_{ij} \sim \mathrm{Geom}(z_{ij})$. Let $N := \sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j$ denote the total sum.*

(i) *There exists some absolute constant $\gamma > 0$ such that*

$$N^{-\gamma(m+n)} e^{g(Z)} \le |\mathcal{M}(\mathbf{r},\mathbf{c})| \le e^{g(Z)}.$$

(ii) *Conditioned on being in $\mathcal{M}(\mathbf{r},\mathbf{c})$, table Y is uniform on $\mathcal{M}(\mathbf{r},\mathbf{c})$.*

(iii) *For the constant $\gamma > 0$ in (i), we have*

$$\mathbb{P}\big(Y \in \mathcal{M}(\mathbf{r},\mathbf{c})\big) = e^{-g(Z)}|\mathcal{M}(\mathbf{r},\mathbf{c})| \ge N^{-\gamma(m+n)}.$$

In other words, (ii) and (iii) of the above theorem says that the *geometric matrix Y* with mean given by the typical table Z emulates the uniform random table X in $\mathcal{M}(\mathbf{r},\mathbf{c})$ with probability at least $N^{-\gamma(m+n)}$. Hence on very rare events, we can 'transfer' some of the properties of this geometric matrix Y to the uniform random contingency table X.

Now we are ready to prove the key lemma.

Proof of Lemma 4.2. Let $\mathcal{B}_1,\dots,\mathcal{B}_k$ be (not necessarily distinct) blocks in $\mathcal{M}(\mathbf{r},\mathbf{c})$ with $|\mathcal{B}_1| \le \dots \le |\mathcal{B}_k|$. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}(\mathbf{r},\mathbf{c})$, and let $Y = (Y_J)$ denote the random matrix of independent entries where $Y_{ij} \sim \mathrm{Geom}(z_{ij})$. Observe that we can choose a subset $\mathcal{I} \subseteq \mathcal{B}_1 \times \dots \times \mathcal{B}_k$ such that $|\mathcal{I}| \ge \lfloor |\mathcal{B}_1|/k \rfloor$ and every matrix coordinate (i,j) appears in at most one element of $\mathcal{I}$. Indeed, in the worst case when $\mathcal{B}_1 = \dots = \mathcal{B}_k$, we may partition $\mathcal{B}_1$ into sub-blocks so that we have at least $\lfloor |\mathcal{B}_1|/k \rfloor$ sub-blocks of size k. We can then turn each sub-block into an element of $\mathcal{I}$. A similar argument holds for the general case.

Fix measurable sets $A \subseteq [0,\infty)$ and $\mathcal{A} \subseteq \mathbb{R}_+^{mn}$. For a $m \times n$ matrix $W = (w_{ij})$ and $I \in \mathcal{B}_1 \times \dots \times \mathcal{B}_k$, denote

$$W^I := \prod_{\ell=1}^{k} w_{I_\ell}, \qquad S(W) := \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \mathbf{1}_{\{W^I \in A\}}.$$

Note that by the exchangeability of the entries of X and Y in each block of $\mathcal{M}(\mathbf{r},\mathbf{c})$, variables X^I and Y^I have the same distribution for all I. In particular, we have:

$$\mathbb{P}(X^I \in A) = \mathbb{E}\big[S(X)\big].$$

Moreover, since Y_{ij} are independent and since every two elements of $\mathscr{I}$ have non-overlapping coordinates, it follows that Y^I are also independent.

Now note that from Theorem 4.3 (ii) and (iii), we have:

$$\mathbb{P}(Y \in \mathscr{A}) \geq \mathbb{P}\big(Y \in \mathscr{A} \mid Y \in \mathscr{M}(\mathbf{r}, \mathbf{c})\big) N^{-\gamma(m+n)} = \mathbb{P}(X \in \mathscr{A}) N^{-\gamma(m+n)}.$$

Also, by the Azuma–Hoeffding inequality, for every fixed $I \in \mathscr{I}$, we have:

$$\mathbb{P}\Big(\big|S(Y) - \mathbb{P}(Y^I \in A)\big| \geq t\Big) \leq 2\exp\left(-\frac{|\mathscr{I}| t^2}{2}\right) \leq 2\exp\left(-\frac{|\mathscr{B}_1| t^2}{2k}\right).$$

Hence, by conditioning on whether $|S(X) - \mathbb{P}(Y_I \in A)|$ is small or large, we get

$$\begin{aligned}
|\mathbb{P}(X^I \in A) - \mathbb{P}(Y^I \in A)| &= \left|\mathbb{E}\big[S(X)\big] - \mathbb{P}(Y^I \in A)\right| \leq \mathbb{E}\left[\big|S(X) - \mathbb{P}(Y^I \in A)\big|\right] \\
&\leq t\,\mathbb{P}\left(\big|S(X) - \mathbb{P}(Y^I \in A)\big| \leq t\right) + 2\mathbb{P}\left(\big|S(X) - \mathbb{P}(Y^I \in A)\big| > t\right) \\
&\leq t + 4N^{\gamma(m+n)}\exp\left(-\frac{|\mathscr{B}_1| t^2}{2k}\right).
\end{aligned}$$

Since $A \subseteq [0, \infty)$ is arbitrary, by absorbing the factor of 4 into γ, we obtain the result. $\qquad\square$

Remark 4.4. Following the arguments in [Bar09, Bar10b], it is not hard to see that a higher dimensional analogue of Theorem 4.3 holds. Namely, replace $\mathscr{M}(\mathbf{r}, \mathbf{c})$ with $\mathscr{M}(\mathbf{r}_1, \ldots, \mathbf{r}_d)$ and $m + n$ with $n_1 + \ldots + n_d$ in the theorem. Of course, the constant $\gamma = \gamma(d) > 0$ then depends on d. Then a similar argument will show that a higher dimensional analogue of Lemma 4.2 also holds. Hence most of our main results should hold in higher dimensions. We do not justify this claim in the present paper.

5. Phase transition in the typical table and rate of convergence

Recall the definition of the typical table Z for $\mathscr{M}(\mathbf{r}, \mathbf{c})$ given in the introduction. Namely, Z is the unique maximizer of the function g defined at (4.1) on the transportation polytope $\mathscr{P}(\mathbf{r}, \mathbf{c})$. Note that $\mathscr{P}(\mathbf{r}, \mathbf{c})$ is defined by the intersection of the hyperplanes in $\mathbb{R}_+^{mn}$ given by

$$h_{i\bullet} := \left(\sum_{k=1}^m x_{ik}\right) - r_i = 0 \qquad \text{for all } 1 \leq i \leq m,$$

$$h_{\bullet j} := \left(\sum_{k=1}^n x_{kj}\right) - c_j = 0 \qquad \text{for all } 1 \leq j \leq n.$$

Also, given that the margins $\mathbf{r} = (r_1, \ldots, r_m)$ and $\mathbf{c} = (c_1, \ldots, c_m)$ have strictly positive coordinates (which is true in our setting), the transportation polytope $\mathscr{P}(\mathbf{r}, \mathbf{c})$ contains a matrix $W = (w_{ij})$ whose entries are all strictly positive. For instance, the expectation under the Fisher-Yates distribution (1.2) has the form $W = (w_{ij})$ with $w_{ij} = r_i c_j / N > 0$, where $N = r_1 + \ldots + r_m$. It follows that the relative interior of $\mathscr{P}(\mathbf{r}, \mathbf{c})$ contains an ε-ball around W for some $\varepsilon > 0$. This allows us to use Lagrange's method to maximize g over the polytope $\mathscr{P}(\mathbf{r}, \mathbf{c})$.

Note that the gradient $\nabla h_{i\bullet}$ is the $m \times n$ matrix $E_{i\bullet}$, which has 1's in the i-th row and 0's elsewhere. Similarly, $\nabla h_{j\bullet}$ is the $m \times n$ matrix $E_{\bullet j}$, which has 1's in the j-th column and 0's elsewhere. On the other hand, it is easy to see that the gradient ∇g of the objective function g defined at (4.1) is given by

$$\nabla g = \big(\log(1 + 1/x_{ij})\big)_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Hence by the multivariate Lagrange's method, when evaluated at the typical table $Z = (z_{ij})$, ∇g must be in the non-negative span of $\nabla h_{i\bullet}$'s and $\nabla h_{\bullet j}$'s. This gives that there exists some non-negative constants

$\lambda_1, \dots, \lambda_m$ and $\mu_1, \dots, \mu_n$ such that

$$\log\left(1 + x_{ij}^{-1}\right) = \lambda_i + \mu_j \qquad \text{for all } 1 \le i \le m, 1 \le j \le n,$$

or equivalently,

(5.1) $$z_{ij} = \frac{1}{\exp(\lambda_i + \mu_j) - 1} \qquad \text{for all } 1 \le i \le m, 1 \le j \le n.$$

Now we consider $\mathcal{M}(\mathbf{r}, \mathbf{c}) = \mathcal{M}_{n,\delta}(B, C)$, where the margins $\mathbf{r}$ and $\mathbf{c}$ are given at (2.1). By symmetry, there exist some constants $\alpha, \beta > 0$, possibly depending on all parameters, such that

$$\log\left(1 + z_{ij}^{-1}\right) = \begin{cases} 2\alpha & \text{if } 1 \le i, j \le \lfloor n^\delta \rfloor, \\ 2\beta & \text{if } \lfloor n^\delta \rfloor < i, j \le \lfloor n^\delta \rfloor + n, \\ \alpha + \beta & \text{otherwise}. \end{cases}$$

Furthermore, denote $P = P_n = \exp(\alpha)$ and $Q = Q_n = \exp(\beta)$. Then (5.1) gives

(5.2) $$z_{11} = \frac{1}{P^2 - 1}, \qquad z_{1,n+1} = \frac{1}{PQ - 1}, \qquad z_{n+1,n+1} = \frac{1}{Q^2 - 1}.$$

Note that the margin condition for Z reduces to

(5.3) $$\begin{cases} (\lfloor n^\delta \rfloor / n)\, z_{11} + z_{1,n+1} = BC, \\ (\lfloor n^\delta \rfloor / n)\, z_{1,n+1} + z_{n+1,n+1} = C. \end{cases}$$

For a preliminary analysis for the solution of the equations in (5.3), observe that

(5.4) $$z_{n+1,n+1} \le C, \quad z_{1,n+1} \le BC, \quad |z_{n+1,n+1} - C| = n^{\delta-1} z_{1,n+1} \le BC n^{\delta-1}.$$

In particular, $z_{n+1,n+1} \to C$ as $n \to \infty$.

The main result in this section is the following lemma, which establishes the phase transition of the typical table and the rate of convergence of its entries.

Lemma 5.1. *Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{M}_{n,\delta}(B, C)$, *where* $0 \le \delta < 1$. *Let* $B_c = 1 + \sqrt{1 + 1/C}$ *be as above.*

(i) *If* $B < B_c$, *then*

$$z_{11} = \frac{B^2 C(C+1)}{(B_c - B)(B_c + B - 2)} + O(n^{\delta-1}), \qquad z_{1,n+1} = BC + O(n^{\delta-1}).$$

(ii) *If* $B > B_c$, *then*

$$z_{n+1,n+1} = C + O(n^{\delta-1}), \qquad z_{1,n+1} = B_c C + O(n^{\delta-1}), \qquad n^{\delta-1} z_{11} = C(B - B_c) + O(n^{\delta-1}).$$

where the constants in $O(n^{\delta-1})$ *remain bounded as* $B \to \infty$.

In the following proposition, we show that if $B < B_c$, then the corner entry z_{11} of the typical table for $\mathcal{M}_{n,\delta}(B, C)$ is uniformly bounded in n. We remark that there is a more general result of this type. Namely, [BLSY10, Thm 3.5] states that if the row and columns do not vary much, then the entries of the typical table are uniformly bounded by some constant independent of the size of the table. However, this result gives a sub-optimal lower critical value $B < (1 + \sqrt{1 + 4(1 + 1/C)})/2$. In order to push this threshold up to the desired critical value $B < B_c$, we optimize the proof of [BLSY10, Thm 3.5] for our model.

Proposition 5.2. *In notation above, suppose* $B < B_c$. *Then:*

$$\limsup_{n \to \infty} z_{11} \le \frac{B^2 C(C+1)}{(B_c - B)(B_c + B - 2)} < \infty.$$

Proof. For brevity, denote $w_{11} = z_{11}$, $w_{21} = w_{12} = z_{1,n+1} = z_{n+1,1}$, and $w_{22} = z_{n+1,n+1}$. Then

$$w_{11} = \frac{1}{P^2 - 1}, \quad w_{21} = w_{22} = \frac{1}{PQ - 1}, \quad w_{22} = \frac{1}{Q^2 - 1}.$$

Note that

$$w_{11} w_{22} = \frac{1}{(P^2 - 1)(Q^2 - 1)} \geq \frac{1}{(PQ - 1)^2} = w_{12} w_{21}.$$

Let us show that

$$w_{21}/w_{22} \leq B.$$

Assume otherwise, that $w_{21}/w_{22} > B$. The above inequality gives $w_{11}/w_{12} \geq w_{12}/w_{22} > B$, and we get

$$BC = \lfloor n^\delta \rfloor w_{11}/n + w_{12} > \lfloor n^\delta \rfloor B w_{21}/n + B w_{22} = BC,$$

a contradiction.

By the definition of P and Q, we have:

$$\frac{(w_{22} + 1) w_{12}}{(w_{12} + 1) w_{22}} = Q^2/PQ = Q/P.$$

In order to upper bound Q/P, we consider maximizing the fraction in the left hand side. By (5.4), we know that $\lim_{n \to \infty} w_{22} = C$. Hence we have the following optimization problem:

$$\text{maximize} \quad \frac{(w_{22} + 1) w_{12}}{(w_{12} + 1) w_{22}}$$
$$\text{subject to} \quad w_{22} \leq w_{12} \leq B w_{22} \quad \text{and} \quad w_{22} = C + o(1).$$

It is not hard to see that the objective function is non-decreasing in w_{12} and non-increasing in w_{22}. Hence, as $n \to \infty$, the solution to the above problem approaches the limit $B(C+1)/(BC+1)$. This implies

$$\limsup_{n \to \infty} \frac{Q}{P} \leq \frac{B(C+1)}{BC+1}.$$

Now, since $w_{12} \leq BC$ by (5.4), we have:

$$\liminf_{n \to \infty} P^2 = \liminf_{n \to \infty} PQ/(Q/P) \geq \liminf_{n \to \infty} \frac{(w_{12} + 1) \cdot P}{w_{12} \cdot Q}$$
$$\geq \frac{(BC + 1) \cdot (BC + 1)}{BC \cdot B(C + 1)} = \frac{(BC + 1)^2}{B^2 C(C + 1)}.$$

This implies

$$\limsup_{n \to \infty} w_{11} = \limsup_{n \to \infty} \frac{1}{P^2 - 1} \leq \frac{B^2 C(C + 1)}{2B - B^2 + C^{-1}} = \frac{B^2 C(C + 1)}{(B_c - B)(B_c + B - 2)}.$$

This finishes the proof. $\qquad\qquad\square$

Proof of Lemma 5.1. Suppose $B < B_c$. By Proposition 5.2, we know that z_{11} is uniformly bounded in n. Hence, from the first equation in (5.3), we obtain:

$$|z_{1,n+1} - BC| \leq n^{\delta - 1} z_{11} = O(n^{\delta - 1}).$$

In particular, $\lim_{n \to \infty} z_{1,n+1} = BC$.

For z_{11}, let $P = P_n$ and $Q = Q_n$ be as in (5.2). Recall that $z_{1,n+1} = 1/(PQ - 1)$ and $z_{n+1,n+1} = 1/(Q^2 - 1)$. Also recall that $z_{n+1,n+1} \to C$ as $n \to \infty$ from (5.4). Hence, we have:

$$\lim_{n \to \infty} \frac{1}{PQ - 1} = BC, \qquad \lim_{n \to \infty} \frac{1}{Q^2 - 1} = C.$$

This implies $Q \to q_* := \sqrt{1+1/C}$ and $P \to p_* := (1+1/BC)/\sqrt{1+1/C} > 1$ as $n \to \infty$. It follows that $z_{11} \to 1/(p_*^2 - 1)$ as $n \to \infty$, which is the correct limit that (i) implies.

In order to obtain the rate of convergence of z_{11}, first define a function $h(x) = 1/(x^2 - 1)$. Then, since $z_{11} = 1/(P^2 - 1)$, it suffices to show

$$(5.5) \qquad |h(P) - h(p^*)| = O(n^{\delta-1}).$$

For this end, first note that $h(x)' = -2x/(x^2-1)^2$ and $|h'(x)|$ is a decreasing function in $(1,\infty)$. Hence by the mean value theorem, for every constant $1 < p < p^*$, we have:

$$(5.6) \qquad |h(P) - h(p^*)| \le |h'(p)| \cdot |P - p^*|$$

for all sufficiently large $n \ge 1$.

Next, write

$$(5.7) \qquad |P - p^*| \le \left| P - \frac{1+1/BC}{Q} \right| + (1+1/BC) \left| \frac{1}{Q} - \frac{1}{q^*} \right|.$$

Since $z_{n+1,n+1} = 1/(Q^2 - 1) = h(Q)$ and $C = h(q_*)$, the mean value theorem implies that

$$(5.8) \qquad |z_{n+1,n+1} - C| = |h(Q) - h(Q_*)| \ge |h'(2q_*)| \cdot |Q - q_*|$$

for all sufficiently large $n \ge 1$. Thus (5.4) gives $|Q - q_*| = O(n^{\delta-1})$. Since $Q \to q_*$ as $n \to \infty$, this also implies that the second term in (5.7) is of order $O(n^{\delta-1})$. For the first term, note that

$$\left| P - \frac{1+1/BC}{Q} \right| = \frac{(PQ-1)/BC}{Q} \left| \frac{1}{PQ-1} - BC \right|.$$

Since $z_{1,n+1} = 1/(PQ - 1)$ and both P and Q converge as $n \to \infty$, the above expression is $O(n^{\delta-1})$. Thus $|P - p_*| = O(n^{\delta-1})$, and (5.5) follows from (5.6). This shows (i).

Next, suppose $B > B_c$. To show (ii), we first obtain a lower bound on z_{11}. Note that (5.4) gives $1/(Q^2 - 1) = z_{n+1,n+1} \le C$, so we have

$$Q \ge \sqrt{1+C^{-1}}.$$

Then from the first equation in (5.3) and the fact that $P \ge 1$, we have

$$z_{11} = (n/\lfloor n^\delta \rfloor)(BC - z_{1,n+1}) \ge n^{1-\delta} \left(BC - \frac{1}{Q-1} \right)$$

$$(5.9) \qquad \ge n^{1-\delta} \left(BC - \frac{1}{\sqrt{1+1/C}-1} \right) = n^{1-\delta} C(B - B_c).$$

Now we derive the limits in (ii). First, note that from (5.9), we have

$$\liminf_{n\to\infty} \frac{n^{\delta-1}}{P^2 - 1} = \liminf_{n\to\infty} n^{\delta-1} z_{11} \ge C(B - B_c) > 0.$$

Since $\delta < 1$, we must have $\lim_{n\to\infty} P_n = 1$. Moreover, $\lim_{n\to\infty} Q_n = q_* = \sqrt{1+1/C}$ by (5.4). We conclude that $\lim_{n\to\infty} z_{1,n+1} = 1/(q_* - 1) = B_c C$.

Finally, we derive the rate of convergence. First, using (5.4) and the limit of $z_{1,n+1}$ we have just shown, we have

$$(5.10) \qquad |z_{n+1,n+1} - C| \le 2B_c C n^{\delta-1}$$

for all sufficiently large $n \ge 1$. Also, from (5.9), we have

$$(5.11) \qquad 0 \le P - 1 \le \frac{n^{\delta-1}}{C(B - B_c)} \quad \text{for all } n \ge 1.$$

Using the notation $q_* = \sqrt{1 + 1/C}$ as above, we can write

$$|z_{1,n+1} - B_c C| = \left| z_{1,n+1} - \frac{1}{q_* - 1} \right| \le \frac{Q(P-1)}{(PQ-1)(Q-1)} + \left| \frac{1}{Q-1} - \frac{1}{q_* - 1} \right|.$$

By (5.11), the first term in the right hand side is $O(n^{\delta-1}/(B-B_c))$, where the constant does not depend on B. On the other hand, note that the inequality (5.8) still holds for $B > B_c$. Hence (5.10) implies that the second term is of order $O(n^{\delta-1})$, where the constant is independent in B. Lastly, it then follows that

$$|n^{\delta-1} z_{11} - C(B - B_c)| = |B_c C - z_{1,n+1}| = O(n^{\delta-1}),$$

where the constant in the last $O(n^{\delta-1})$ does not grow in B. This proves (ii). $\qquad\square$

6. Convergence in total variation distance and the proof of Theorem 2.1

In this section we prove Theorem 2.1. We first derive the following result from Lemma 4.2, which states that each entry of the random contingency table $X \in \mathcal{M}_{n,\delta}(B,C)$ has a geometric limiting distribution with mean dictated by the corresponding entry in the typical table.

Theorem 6.1. *Let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B,C)$ uniformly at random. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B,C)$ and let $Y = (Y_{ij})$ denote the random matrix of independent entries with $Y_{ij} \sim \mathrm{Geom}(z_{ij})$. Fix integers $k \ge 1$ and $1 \le i_r, j_r \le n + \lfloor n^\delta \rfloor$, $1 \le r \le k$. Then for every $\varepsilon > 0$, we have:*

$$d_{TV}\left(\prod_{r=1}^{k} X_{i_r, j_r}, \prod_{r=1}^{k} Y_{i_r, j_r} \right) \le n^{-\eta(\delta)+\varepsilon}$$

for all sufficiently large $n \ge 1$, where

(6.1)
$$\eta(\delta) = \begin{cases} 1/2 & \text{if all } (i_r, j_r)\text{'s are contained in the bottom right block,} \\ \delta - 1/2 & \text{if some } (i_r, j_r) \text{ is contained in the top left block,} \\ \delta/2 & \text{otherwise.} \end{cases}$$

Proof. Fix $\varepsilon > 0$. Note that each (i_r, j_r) is contained one of the four blocks of $\mathcal{M}_{n,\delta}(B,C)$. Suppose that all of the indices are contained in the bottom right block $\mathscr{B}_{BR} := \{\lfloor n^\delta \rfloor + 1, \lfloor n^\delta \rfloor + n\}^2$, which has size n^2. Then we apply Lemma 4.2 with $t = (1/2)n^{-1/2+\varepsilon}$. Since $N \le Cn^2 + BCn^{1+\delta}$ and $0 \le \delta < 1$, this gives:

$$d_{TV}\left(\prod_{r=1}^{k} X_{i_r, j_r}, \prod_{r=1}^{k} Y_{i_r, j_r} \right) \le \frac{1}{2} n^{-1/2+\varepsilon} + \exp(c\, n \log n) \exp(-c\, n^{1+2\varepsilon})$$

$$\le \frac{1}{2} n^{-1/2+\varepsilon} + \exp(-c' n^{1+2\varepsilon}),$$

for some constants $c, c' > 0$. The equation in the theorem follows from here. Note that the top left block has size $(\lfloor n^\delta \rfloor)^2$ and the top right and bottom left blocks have size $n \lfloor n^\delta \rfloor$. By applying Lemma 4.2 for $t = (1/2)n^{-(\delta-1/2)+\varepsilon}$ and $t = (1/2)n^{-(\delta/2)+\varepsilon}$ for the other cases, respectively, the theorem now follows from a similar argument. $\qquad\square$

Remark 6.2. For marginal distribution of single entry of $X \in \mathcal{M}_{n,\delta}(B,C)$, our Theorem 6.1 implies:

$$\begin{cases} d_{TV}(X_{n+1,n+1}, Y_{n+1,n+1}) \le n^{-1/2+\varepsilon} \\ d_{TV}(X_{1,n+1}, Y_{1,n+1}) \le n^{-(\delta/2)+\varepsilon} \\ d_{TV}(X_{11}, Y_{11}) \le n^{1/2-\delta+\varepsilon}, \end{cases}$$

for all $\varepsilon > 0$ and sufficiently large $n \ge 1$.

Next, we give a simple upper bound on the total variation distance between two geometric distributions in terms of the difference of their mean.

Proposition 6.3. *For all $\lambda, \lambda' > 0$, we have:*

$$d_{TV}(\mathrm{Geom}(\lambda), \mathrm{Geom}(\lambda')) \le 2|\lambda - \lambda'| \min\left(\frac{1+\lambda}{1+\lambda'}, \frac{1+\lambda'}{1+\lambda}\right).$$

Proof. Suppose $0 \le p < q \le 1$. Then for every $k \ge 0$,

$$(1-p)p^k - (1-q)q^k = (p^k - q^k) - (p^{k+1} - q^{k+1})$$
$$= (p-q)\left[(p^{k-1} + p^{k-2}q + \ldots + pq^{k-1} + q^k) - (p^k + p^{k-1}q + \ldots + pq^k + q^{k+1})\right],$$

so this gives

$$\sum_{k=0}^{\infty} |(1-p)p^k - (1-q)q^k| \le 2|p-q| \sum_{k=0}^{\infty} (k+1)q^k \le \frac{2|p-q|}{(1-q)^2}.$$

Fix a constant $\lambda > 0$ and let $X \sim \mathrm{Geom}(\lambda)$. Then:

$$\mathbb{P}(X = k) = \left(\frac{1}{1+\lambda}\right)\left(\frac{\lambda}{1+\lambda}\right)^k, \qquad k \in \mathbb{N}.$$

For $X' \sim \mathrm{Geom}(\lambda')$ with $\lambda' > 0$, we have:

$$2\,d_{TV}\big(\mathrm{Geom}(\lambda), \mathrm{Geom}(\lambda')\big) = \sum_{k=0}^{\infty} |\mathbb{P}(X = k) - \mathbb{P}(X' = k)| \le 2\left|\frac{1}{1+\lambda} - \frac{1}{1+\lambda'}\right|(1+\lambda)^2 = 2|\lambda - \lambda'|\frac{1+\lambda}{1+\lambda'}.$$

Then the result follows by changing the roles of λ and λ'. $\qquad\square$

Proof of Theorem 2.1. In order to show (i), first note that by Proposition 6.3 and (5.4), we get:

$$d_{TV}\big(\mathrm{Geom}(z_{n+1,n+1}), \mathrm{Geom}(C)\big) \le |z_{n+1,n+1} - C|\frac{1+C}{1+z_{n+1}} \le n^{\delta-1}BC(1+C).$$

Then the bound on the total deviation distance between $X_{n+1,n+1}$ and $\mathrm{Geom}(C)$ follows from Theorem 6.1 (see also Remark 6.2) and the triangle inequality. This shows (i).

Next, suppose $B < B_c$. Then, by Theorem 6.1 (ii) (see also Remark 6.2) and Proposition 6.3, for all $\varepsilon > 0$ and sufficiently large $n \ge 1$, we have:

$$d_{TV}\big(X_{1,n+1}, \mathrm{Geom}(BC)\big) \le d_{TV}\big(X_{1,n+1}, \mathrm{Geom}(z_{1,n+1})\big) + d_{TV}\big(z_{1,n+1}, \mathrm{Geom}(BC)\big)$$
$$\le d_{TV}\big(X_{1,n+1}, \mathrm{Geom}(z_{1,n+1})\big) + d_{TV}\big(z_{1,n+1}, \mathrm{Geom}(BC)\big)$$
$$\le n^{-(\delta/2)+\varepsilon} + 2|z_{1,n+1} - BC|(1+z_{1,n+1})(1+BC).$$

Therefore, the second term is of order $O(n^{\delta-1})$ by Lemma 5.1 and (5.4). A similar argument can be used to prove the rest of the theorem, except for the case of X_{11} at the supercritical regime.

For the last case, suppose $B > B_c$ consider X_{11}. Denote $\sigma := C(B - B_c)$ and $\xi_n := n^{1-\delta}\sigma - z_{11}$ for each $n \ge 1$. Note that $\xi_n = O(1)$ by Lemma 5.1 (ii). Then for all $\varepsilon > 0$ and sufficiently large $n \ge 1$, Theorem 6.1 implies:

$$d_{TV}\big(X_{11}, \mathrm{Geom}(\sigma n^{1-\delta} + \xi_n)\big) = d_{TV}\big(X_{11}, \mathrm{Geom}(z_{11})\big) \le n^{1/2-\delta+\varepsilon}.$$

This finishes the proof. $\qquad\square$

7. Convergence of first moments and the proof of Theorem 2.2

In this section, we provide convergence of first moments of the entries in the uniform contingency table $X \in \mathcal{M}_{n,\delta}(B,C)$, and prove Theorem 2.2. Depending on the block that the entries belong to, the convergence may only hold under truncation for general $0 < \delta < 1$. For $1/2 < \delta < 1$, we can get rid of all truncation by using Theorem 2.1.

We first prove some useful facts that allow us to transfer convergence in total variation distance into L_1 convergence under a certain uniform boundedness condition. We write $a \wedge b = \min(a,b)$ for all $a,b \in \mathbb{R}$.

Proposition 7.1. *Let* $(Y_n)_{n\geq 0}$ *be a sequence of random variables taking values from* $\mathbb{N}$ *with finite mean. Suppose that there exists some constant* $c > 0$ *such that* $\mathbb{P}(Y_n \geq k) \leq \exp(-ck)$ *for all* $n, k \geq 0$. *Then for every sequence* $M_n \to \infty$, *we have:*

$$|\mathbb{E}[Y_n \wedge M_n] - \mathbb{E}[Y_n]| = O\big(\exp(-cM_n)\big).$$

Proof. Fix $M > 0$. First, observe that

$$\sum_{k>M} \mathbb{P}(Y_n \geq k) = \sum_{k>M} \sum_{j\geq k} \mathbb{P}(Y_n = j) = \sum_{j>M} \sum_{M<k\leq j} \mathbb{P}(Y_n = j)$$
$$= \sum_{j>M} (j - M)\mathbb{P}(X_n = j) = \mathbb{E}\big[Y_n - Y_n \wedge M\big].$$

By the exponential tail bound on Y_n's this gives

$$\mathbb{E}[Y_n - Y_n \wedge M] \leq \sum_{k>M} e^{-ck} = \frac{e^{-c(M+1)}}{1 - e^{-c}}.$$

Letting $M := M_n \to \infty$ as $n \to \infty$ implies the result. $\square$

Proposition 7.2. *Let* $(X_n)_{n\geq 0}$ *and* $(Y_n)_{n\geq 0}$ *be sequences of random variables taking values from* $\mathbb{N}$ *with the following properties:*

(i) *There exists some constant* $c > 0$ *such that* $\mathbb{P}(Y_n \geq k) \leq \exp(-ck)$ *for all* $n, k \geq 0$.
(ii) $d_{TV}(X_n, Y_n) = O(n^{-\beta})$ *for some* $\beta > 0$.

Then for every sequence $M_n \to \infty$, *we have:*

$$\left|\mathbb{E}[X_n \wedge M_n] - \mathbb{E}[Y_n]\right| = O\big(M_n n^{-\beta}\big).$$

Proof. Fix $M > 0$. Write

$$\mathbb{E}[X_n \wedge M] - \mathbb{E}[Y_n \wedge M] = \sum_{k=0}^{M} k\delta_{k,n},$$

where $\delta_{k,n} = \mathbb{P}(X_n = k) - \mathbb{P}(Y_n = k)$. Note that

$$\sum_{k=0}^{M} |\delta_{k,n}| \leq \sum_{k=0}^{\infty} |\delta_{k,n}| = 2 d_{TV}(X_n, Y_n).$$

Hence we have

$$|\mathbb{E}[X_n \wedge M] - \mathbb{E}[Y_n \wedge M]| \leq \left|\sum_{k=0}^{M} k\delta_{k,n}\right| \leq M \sum_{k=0}^{M} |\delta_{k,n}| \leq 2M d_{TV}(X_n, Y_n).$$

By the assumption (ii), this implies

$$|\mathbb{E}[X_n \wedge M_n] - \mathbb{E}[Y_n \wedge M_n]| = O(M_n n^{-\beta}).$$

Now the result follows from the triangle inequality and Proposition 7.1. $\square$

In the rest of this section, we let $X = (X_{ij})$ denote the random contingency table sampled from $\mathcal{M}_{n,\delta}(B,C)$ uniformly at random. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B,C)$.

As an immediate application of Lemma 5.1, Theorem 2.1, and Propositions 7.1 and 7.2, we show that the first moments of the entries in X converge to the corresponding entries in Z under truncation.

Proposition 7.3. *Let $B, C > 0$ and let $B_c = 1 + \sqrt{1 + 1/C}$ as above. Fix $\alpha, \varepsilon > 0$.*

(i) *For all $0 \le \delta < 1$, we have:*
$$|\mathbb{E}\left[X_{n+1,n+1} \wedge n^\alpha\right] - z_{n+1,n+1}| = O\big(n^{\alpha - 1/2 + \varepsilon}\big).$$

(ii) *For all $0 < \delta < 1$, we have:*
$$|\mathbb{E}\left[X_{1,n+1} \wedge n^\alpha\right] - z_{1,n+1}| = O\big(n^{\alpha - (\delta/2) + \varepsilon}\big).$$

(iii) *For all $1/2 < \delta < 1$, we have:*
$$\begin{cases} |\mathbb{E}\left[X_{11} \wedge n^\alpha\right] - z_{11}| = O(n^{\alpha + 1/2 - \delta + \varepsilon}) & \text{if } B < B_c, \\ \left|\mathbb{E}\left[(n^{\delta-1} X_{11}) \wedge n^\alpha\right] - n^{\delta-1} z_{11}\right| = O(n^{\alpha + 1/2 - \delta + \varepsilon}) & \text{if } B > B_c. \end{cases}$$

Proof. We first show (i). Let $Y = (Y_{ij})$ be the matrix of independent entries where Y_{ij} has geometric distribution with mean z_{ij}. Recall that $z_{n+1,n+1} \le C$ from (5.4). Hence $Y_{n+1,n+1}$ has exponential tail bound independent of n. Since $\mathbb{E}[Y_{n+1,n+1}] = z_{n+1,n+1}$, (i) follows from Remark 6.2 and Proposition 7.2. A similar argument gives (ii) and (iii) for all $B < B_c$. Lastly, suppose $B > B_c$ and $1/2 < \delta < 1$. Note that by Lemma 5.1, we have:
$$\lim_{n \to \infty} n^{\delta-1} z_{11} = C(B - B_c).$$

Thus $n^{\delta-1} Y_{11}$ has exponential tail bound independent of n. Hence (iii) for $B > B_c$ also follows from Remark 6.2 and Proposition 7.2. $\square$

In the following proposition, we show that in many cases, we can get rid of the truncation in Proposition 7.3. For this purpose, we work with the following 'error table' $\overline{X} = (\overline{X}_{ij}) := X - Z$. If we denote $r = \lfloor n^\delta \rfloor$, then the margin condition for the tables in $\mathcal{M}_{n,\delta}(B,C)$ give:

$$(7.1) \quad \begin{cases} (\overline{X}_{11} + \overline{X}_{12} + \ldots + \overline{X}_{1r}) + (\overline{X}_{1,r+1} + \overline{X}_{1,r+2} + \ldots + \overline{X}_{1,r+n}) & = 0, \\ (\overline{X}_{n+1,1} + \overline{X}_{n+1,2} + \ldots + \overline{X}_{n+1,r}) + (\overline{X}_{n+1,r+1} + \overline{X}_{n+1,r+2} + \ldots + \overline{X}_{n+1,r+n}) & = 0. \end{cases}$$

Proposition 7.4. *Fix $B, C > 0$ and $0 \le \delta < 1$. Let $B_c = 1 + \sqrt{1 + 1/C}$. Then for all $\varepsilon > 0$, we have:*

$$\begin{cases} \left|\mathbb{E}\left[\overline{X}_{n+1,n+1}\right]\right| = O(n^{\delta-1}) & \text{for all } 0 \le \delta < 1, \\ n^{\delta-1} |\mathbb{E}[X_{11}]| + \left|\mathbb{E}\left[\overline{X}_{1,n+1}\right]\right| = O(n^{\delta-1} + n^{-(\delta/2)+\varepsilon}) & \text{for all } B < B_c \text{ and } 0 < \delta < 1, \\ n^{\delta-1} \left|\mathbb{E}\left[\overline{X}_{11}\right]\right| + \left|\mathbb{E}\left[\overline{X}_{1,n+1}\right]\right| = O(n^{(1/2)-\delta+\varepsilon}) & \text{for all } B > B_c \text{ and } 1/2 < \delta < 1. \end{cases}$$

Proof. Taking expectation for the equations in (7.1) and using symmetry of entries in each block, we obtain:

$$(7.2) \quad \begin{cases} (r/n)\mathbb{E}\left[\overline{X}_{11}\right] + \mathbb{E}\left[\overline{X}_{1,n+1}\right] = 0, \\ (r/n)\mathbb{E}\left[\overline{X}_{n+1,1}\right] + \mathbb{E}\left[\overline{X}_{n+1,n+1}\right] = 0. \end{cases}$$

Note that $\mathbb{E}[X_{n+1,1}] \le BC$ and $z_{n+1,1} \le BC$, so $|\mathbb{E}[\overline{X}_{n+1,1}]| \le 2BC$. Hence the second equation in (7.2) gives the first equation in the proposition.

In other to show the second and the third equations, suppose $0 < \delta < 1$. Fix $\alpha, \varepsilon > 0$ and denote $x^+ = \max(x, 0)$. Note that

$$(7.3) \quad \mathbb{E}[\overline{X}_{1,n+1}] = \mathbb{E}\left[X_{1,n+1} \wedge n^\alpha - z_{1,n+1}\right] + \mathbb{E}\left[\left(X_{1,n+1} - n^\alpha\right)^+\right].$$

Also, by Proposition 7.3, we have:

$$(7.4) \qquad |\mathbb{E}\left[X_{1,n+1} \wedge n^\alpha - z_{1,n+1}\right]| = O\big(n^{\alpha-(\delta/2)+\varepsilon}\big).$$

Suppose $B < B_c$. From the first row sum condition for X, we have:

$$(7.5) \qquad (r/n)\mathbb{E}[X_{11}] + \mathbb{E}\left[X_{1,n+1} \wedge n^\alpha - z_{1,n+1}\right] + \mathbb{E}\left[\left(X_{1,n+1} - n^\alpha\right)^+\right] = \frac{\lfloor BCn \rfloor}{n} - z_{1,n+1}.$$

The right hand side is of order $O(n^{\delta-1})$ by Lemma 5.1. Since the first and the last terms in the left hand side are nonnegative, this shows:

$$n^{\delta-1}\mathbb{E}[X_{11}] = O\big(n^{\delta-1} + n^{\alpha-(\delta/2)+\varepsilon}\big), \quad \mathbb{E}\left[\left(X_{1,n+1} - n^\alpha\right)^+\right] = O\big(n^{\delta-1} + n^{\alpha-(\delta/2)+\varepsilon}\big).$$

Furthermore, from (7.3) and (7.4), we have:

$$\left|\mathbb{E}[\overline{X}_{1,n+1}]\right| = O(n^{\delta-1} + n^{\alpha-(\delta/2)+\varepsilon}).$$

Since $\alpha, \varepsilon > 0$ were arbitrary, letting $\alpha \to 0$ gives the second equation in the proposition.

It remains to show the last equation the proposition. Suppose $B > B_c$ and $1/2 < \delta < 1$. We rewrite (7.5) as

$$\frac{\lfloor BCn \rfloor}{n} - z_{1,n+1} - (r/n)z_{11} = \mathbb{E}\left[(r/n)X_{11} \wedge n^\alpha - (r/n)z_{11}\right] + \mathbb{E}\left[X_{1,n+1} \wedge n^\alpha - z_{1,n+1}\right]$$
$$+ \mathbb{E}\left[\left((r/n)X_{11} - n^\alpha\right)^+\right] + \mathbb{E}\left[\left(X_{1,n+1} - n^\alpha\right)^+\right].$$

By Lemma 5.1, the left hand side is of order $O(n^{\delta-1})$. The first term in the right hand side is of order $O(n^{\alpha+(1/2)-\delta+\varepsilon})$ by Proposition 7.3. Noting the bound in (7.4) and that the last two terms in the right hand side are nonnegative, we deduce

$$\mathbb{E}\left[\left((r/n)X_{11} - n^\alpha\right)^+\right] = O\big(n^{\delta-1} + n^{\alpha+(1/2)-\delta+\varepsilon}\big), \quad \mathbb{E}\left[\left(X_{1,n+1} - n^\alpha\right)^+\right] = O\big(n^{\delta-1} + n^{\alpha+(1/2)-\delta+\varepsilon}\big).$$

This gives

$$n^{\delta-1}\mathbb{E}\left[\overline{X}_{11}\right] = O\big(n^{\delta-1} + n^{\alpha+(1/2)-\delta+\varepsilon}\big), \qquad \mathbb{E}\left[\overline{X}_{1,n+1}\right] = O\big(n^{\delta-1} + n^{\alpha+(1/2)-\delta+\varepsilon}\big).$$

Since $\alpha, \varepsilon > 0$ are arbitrary, letting $\alpha \to 0$ proves the last equation in the proposition and completes the proof. $\qquad\square$

We can now derive Theorem 2.2 from Lemma 5.1, Proposition 7.3 and Proposition 7.4.

Proof of Theorem 2.2. Let $\overline{X} = (\overline{X}_{ij}) = X - Z$ denote the error table as before. From (5.4) and the first equation in Proposition 7.4, we have

$$|\mathbb{E}[X_{n+1,n+1}] - C| \le |\mathbb{E}[\overline{X}_{n+1,n+1}]| + |z_{n+1,n+1} - C| = O\big(n^{\delta-1}\big),$$

which holds for all $0 \le \delta < 1$ and $B, C > 0$. This shows (i).

Next, we show (ii). Fix $\varepsilon > 0$. Suppose $0 < \delta < 1$ and $B < B_c$. Then by Proposition 7.4 and Lemma 5.1, we have

$$|\mathbb{E}[X_{1,n+1}] - BC| \le |\mathbb{E}[\overline{X}_{1,n+1}]| + |z_{1,n+1} - BC| = O\big(n^{\delta-1} + n^{-(\delta/2)+\varepsilon}\big).$$

Moreover, if $1/2 < \delta < 1$ and $B > B_c$, then similarly we have

$$|\mathbb{E}[X_{1,n+1}] - B_cC| \le |\mathbb{E}[\overline{X}_{1,n+1}]| + |z_{1,n+1} - B_cC| = O\big(n^{\delta-1} + n^{(1/2)-\delta+\varepsilon}\big).$$

Furthermore, if $0 < \delta < 1$ and $B > B_c$, then for every fixed $\alpha > 0$, our Lemma 5.1 and Proposition 7.3 give:

$$|\mathbb{E}[X_{1,n+1} \wedge n^\alpha] - B_cC| \le |\mathbb{E}[X_{1,n+1} \wedge n^\alpha] - z_{1,n+1}| + |z_{1,n+1} - B_cC| = O\big(n^{\alpha-(\delta/2)+\varepsilon} + n^{\delta-1}\big).$$

This shows (ii).

Finally, we obtain (iii) as follows. Fix $\varepsilon > 0$. The first equation for $B < B_c$ follows from the second equation in Proposition 7.4. If we further assume that $1/2 < \delta < 1$, then by Proposition 7.3 and Lemma 5.1, we have:

$$\left| \mathbb{E}[X_{11} \wedge n^\alpha] - \frac{B^2(1+C)}{(B_c - B)(B + B_c - 2)} \right| \le \left| z_{11} - \frac{B^2(1+C)}{(B_c - B)(B + B_c - 2)} \right| +$$
$$+ |\mathbb{E}[X_{11} \wedge n^\alpha] - z_{1,n+1}| = O\big(n^{\delta-1} + n^{\alpha + (1/2) - \delta + \varepsilon}\big).$$

Denote $r = \lfloor n^\delta \rfloor$. For $1/2 < \delta < 1$ and $B > B_c$, by Proposition 7.4 and Lemma 5.1 we have:

$$|(r/n)\mathbb{E}[X_{11}] - C(B - B_c)| \le |(r/n)\mathbb{E}[X_{11}] - (r/n)z_{11}| + |(r/n)z_{11} - C(B - B_c)(r/n)|$$
$$= O\big(n^{\delta-1} + n^{\alpha + (1/2) - \delta + \varepsilon}\big).$$

It remains to prove (iii) for $0 < \delta < 1$ and $B > B_c$. Fix $\alpha > 0$. Note that

$$\frac{\lfloor BCn \rfloor}{n} - B_c C = \frac{\lfloor n^\delta \rfloor}{n} \mathbb{E}[X_{11}] + \mathbb{E}[(X_{1,n+1} - n^\alpha)^+] + (z_{1,n+1} - B_c C) + \mathbb{E}[X_{1,n+1} \wedge n^\alpha - z_{1,n+1}].$$

By Lemma 5.1 and Proposition 7.3 (ii), the last two terms in the right hand side are of order $O(n^{\delta-1})$ and $O(n^{\alpha - (\delta/2) + \varepsilon})$, respectively. Hence the difference between the left hand side and the sum of the first two terms in the right hand side must be of order $O(n^{\delta-1} + n^{\alpha - (\delta/2) + \varepsilon})$. This completes the proof of the theorem. $\square$

8. CONVERGENCE OF HIGHER MOMENTS AND LAW OF LARGE NUMBERS

In this section, we prove Theorem 2.3. Our argument is based on bounding the joint second moments of the entries in the uniform contingency table $X \in \mathcal{M}_{n,\delta}(B, C)$. We begin by generalizing Proposition 7.3 into higher moments.

Proposition 8.1. *Fix $B, C > 0$ and denote $B_c = 1 + \sqrt{1 + 1/C}$. Let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B, C)$ uniformly at random. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B, C)$. Fix integers $k \ge 1$ and $1 \le i_\ell, j_\ell \le n + \lfloor n^\delta \rfloor$, $1 \le \ell \le k$. Let $\eta(\delta)$ be defined as (A1), and denote*

$$m_{11} = \# \text{ of } (i_\ell, j_\ell)\text{'s in the top left block of } \mathcal{M}_{n,\delta}(B, C)$$

Fix $\alpha, \varepsilon > 0$. Then unless both $m_{11} > 0$ and $B = B_c$, for all sufficiently large $n \ge 1$, we have

$$\left| \mathbb{E}\left[\left(\prod_{\ell=1}^{k} X_{i_\ell, j_\ell} \right) \wedge n^{\alpha + (1-\delta)m_{11}\mathbf{1}(B > B_c)} \right] - \prod_{\ell=1}^{k} z_{i_\ell, j_\ell} \right| \le n^{\alpha - \eta(\delta) + \varepsilon}.$$

Proof. Similar to the proof of Proposition 7.3. Details are omitted. $\square$

Proposition 8.2. *Fix $B, C > 0$ and $0 \le \delta < 1$. Let $X = (X_{ij})$ be sampled from $\mathcal{M}_{n,\delta}(B, C)$ uniformly at random. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B, C)$. Denote $(\overline{X}_{ij}) = X - Z$ and $r = \lfloor n^\delta \rfloor$. For $\delta > 1/2$ and all $\varepsilon > 0$, we have:*

$$\mathbb{E}\left[\left(\sum_{\ell=1}^{r} \overline{X}_{1\ell} \right)^2 \right] + \mathbb{E}\left[\left(\sum_{\ell=r+1}^{n+r} \overline{X}_{1\ell} \right)^2 \right] + \left| \mathbb{E}\left[\left(\sum_{\ell=1}^{r} \overline{X}_{1\ell} \right) \left(\sum_{\ell=r+1}^{n+r} \overline{X}_{1\ell} \right) \right] \right| = O\big(n^{(5/2) - \delta + \varepsilon}\big).$$

Furthermore, a similar bound holds for the $(n+1)$-th row for all $0 \le \delta < 1$.

Proof. We only prove the first part of the proposition, for $1/2 < \delta < 1$. A similar argument shows the second part.

Taking square and expectation for the first equation in (7.1) give

$$\sum_{1 \le i, j \le n+r} \mathbb{E}\left[\overline{X}_{1i} \overline{X}_{1j} \right] = 0.$$

Using the symmetry of the entries in each block, we can write the above equation as

$$r\mathbb{E}\left[\overline{X}_{11}^2\right] + n\mathbb{E}\left[\overline{X}_{1,n+1}^2\right] + r(r-1)\mathbb{E}\left[\overline{X}_{11}\overline{X}_{12}\right] + n(n-1)\mathbb{E}\left[\overline{X}_{1,n+1}\overline{X}_{1,n+2}\right] + rn\mathbb{E}\left[\overline{X}_{11}\overline{X}_{1,n+1}\right] = 0.$$

Denote $W_{1i} = r\overline{X}_{1i}/n$ for $1 \le i \le r$. Then we have:

$$\frac{1}{r}\mathbb{E}\left[W_{11}^2\right] + \frac{1}{n}\mathbb{E}\left[\overline{X}_{1,n+1}^2\right] + \left(1 - \frac{1}{r}\right)\mathbb{E}\left[W_{11}W_{12}\right] + \left(1 - \frac{1}{n}\right)\mathbb{E}\left[\overline{X}_{1,n+1}\overline{X}_{1,n+2}\right] + \mathbb{E}\left[W_{11}\overline{X}_{1,n+1}\right] = 0.$$

Now fix $\alpha, \varepsilon > 0$ and denote $x^+ = \max(x, 0)$. Write

$$(8.1) \qquad \mathbb{E}\left[\overline{X}_{1,n+1}\overline{X}_{1,n+2}\right] = \mathbb{E}\left[X_{1,n+1}X_{1,n+2} \wedge n^\alpha - z_{1,n+1}^2\right] + \mathbb{E}\left[\left(X_{1,n+1}X_{1,n+2} - n^\alpha\right)^+\right].$$

Note that by Proposition 8.1, we have

$$(8.2) \qquad \left|\mathbb{E}\left[X_{1,n+1}X_{1,n+2} \wedge n^\alpha - z_{1,n+1}^2\right]\right| = O\left(n^{\alpha-(\delta/2)+\varepsilon}\right).$$

Similarly, we can write

$$\mathbb{E}\left[W_{11}W_{12}\right] = \mathbb{E}\left[(rX_{11}/n)(rX_{12}/n) \wedge n^\alpha - (rz_{11}/n)^2\right] + \mathbb{E}\left[\left((rX_{11}/n)(rX_{12}/n) - n^\alpha\right)^+\right],$$

and Proposition 8.1 gives

$$\left|\mathbb{E}\left[(rX_{11}/n)(rX_{12}/n) \wedge n^\alpha - (rz_{11}/n)^2\right]\right| = O\left(n^{\alpha+(1/2)-\delta+\varepsilon}\right).$$

By truncating the other terms and applying Proposition 8.1 similarly, this gives

$$\frac{1}{r}\mathbb{E}\left[\left((rX_{11}/n)^2 - n^\alpha\right)^+\right] + \frac{1}{n}\mathbb{E}\left[\left(X_{1,n+1}^2 - n^\alpha\right)^+\right] + \left(1 - \frac{1}{r}\right)\mathbb{E}\left[\left((rX_{11}/n)(rX_{12}/n) - n^\alpha\right)^+\right]$$

$$+ \left(1 - \frac{1}{n}\right)\mathbb{E}\left[\left(X_{1,n+1}X_{1,n+2} - n^\alpha\right)^+\right] + \mathbb{E}\left[\left((rX_{11}/n)X_{1,n+1} - n^\alpha\right)^+\right] = O\left(n^{\alpha+(1/2)-\delta+\varepsilon}\right).$$

Since all terms in the left hand side of the above equation are nonnegative, they have order $O(n^{\alpha+(1/2)-\delta+\varepsilon})$ individually. This give bounds on the truncation error for each moments.

From (8.1), (8.2), and the above bound on the truncation error, we deduce that

$$|\mathbb{E}[\overline{X}_{1,n+1}\overline{X}_{1,n+2}]| = O\left(n^{\alpha+(1/2)-\delta+\varepsilon}\right).$$

Since $\alpha, \varepsilon > 0$ were arbitrary, this gives:

$$|\mathbb{E}[\overline{X}_{1,n+1}\overline{X}_{1,n+2}]| = O\left(n^{(1/2)-\delta+\varepsilon}\right) \quad \text{for all } \varepsilon > 0.$$

Similarly, we obtain:

$$|\mathbb{E}[W_{11}^2]| = O\left(n^{(1/2)+\varepsilon}\right), \qquad |\mathbb{E}[\overline{X}_{1,n+1}^2]| = O\left(n^{(3/2)-\delta+\varepsilon}\right),$$

$$|\mathbb{E}[W_{11}W_{12}]| + |\mathbb{E}[X_{11}\overline{X}_{1,n+1}]| = O\left(n^{(1/2)-\delta+\varepsilon}\right).$$

The result now follows by combining these bounds. $\qquad \square$

Proof of Theorem 2.3. We only prove the first part of the theorem, for $1/2 < \delta < 1$. A similar argument shows the second part.

Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B, C)$. Let $S_{n,\delta}(B, C)$ denote the sum of the entries in the first row of the up right block, as defined in Conjecture 3.1. According to Lemma 5.1, as $n \to \infty$,

$$z_{1,n+1} \to \begin{cases} BC & \text{if } B < B_c \\ B_c C & \text{if } B > B_c. \end{cases}$$

Hence by the triangle inequality, it suffices to show that

$$(8.3) \qquad \left|n^{-1}S_{n,\delta}(B, C) - z_{1,n+1}\right| \to 0 \quad \text{as } n \to \infty, \text{ a.s.}$$

Denote $r = \lfloor n^\delta \rfloor$ and $(\overline{X}_{ij}) = X - Z$. Let

$$\overline{S}_{n,\delta}(B,C) := S_{n,\delta}(B,C) - n\,z_{1,n+1} = \overline{X}_{1,r+1} + \overline{X}_{1,r+2} + \ldots + \overline{X}_{1,r+n}.$$

By the Markov inequality, we have:

$$\mathbb{P}\left(\overline{S}_{n,\delta}(B,C) \geq t\right) \leq \frac{1}{t^2}\,\mathbb{E}\left[\left(\sum_{k=r+1}^{n+r} \overline{X}_{1k}\right)^2\right].$$

Hence, by Proposition 8.2, we have that for every $\xi, \varepsilon > 0$ there exists a constant $c > 0$, s.t.

$$\mathbb{P}\left(\overline{S}_{n,\delta}(B,C) \geq n^{1-\xi}\right) \leq c\,n^{(1/2)-\delta+2\xi+\varepsilon}$$

for all sufficiently large $n \geq 1$. Thus, if we choose $0 < \xi < 1/2 - \delta$, then for some $\xi', c' > 0$ we have:

$$\mathbb{P}\left(\overline{S}_{n,\delta}(B,C) \geq n^{1-\xi}\right) \leq c'\,n^{-\xi'}$$

for all sufficiently large $n \geq 1$. This implies that for every sequence $(n_k)_{k \geq 1}$ s.t. $n_k \to \infty$ as $k \to \infty$, we can choose a subsequence $n_{k(r)} \to \infty$ as $r \to \infty$, s.t.

$$\sum_{r=1}^{\infty} \mathbb{P}\left(\frac{\overline{S}_{n_{k(r)},\delta}(B,C)}{n_{k(r)}} \geq (n_{k(r)})^{-\xi}\right) < \infty.$$

By the Borel–Cantelli lemma, it follows that

$$\frac{\overline{S}_{n_{k(r)},\delta}(B,C)}{n_{k(r)}} \to 0 \quad \text{as} \;\; r \to \infty, \text{ a.s.}$$

Therefore, almost surely,

$$\liminf_{n\to\infty} n^{-1}\overline{S}_{n,\delta}(B,C) = \limsup_{n\to\infty} n^{-1}\overline{S}_{n,\delta}(B,C) = 0.$$

This shows (8.3), and completes the proof.   □

9. FINAL REMARKS AND OPEN PROBLEMS

9.1. **Large C limit.** In the limiting case $C \to \infty$ our results give continuous distributions, which can be viewed as results on projections of transportation polytopes $\mathscr{P}(\mathbf{r},\mathbf{c})$, with $\mathbf{r} = \mathbf{c}$ given by (2.1) for $C = 1$. Scaled properly, in the limit the geometric distribution becomes the exponential distribution, so we recover the Theorem 1 in [CDS10] from our Theorem 2.1.

Note that in the continuous case the critical value

$$B_c = \lim_{C\to\infty} 1 + \sqrt{1 + 1/C} = 2.$$

In [DLP19+], the authors tested the conjectures for the larger values of C. Figure 4 is similar to Figure 2 with $C = 9$, with a subcritical $B = 1.8$ and supercritical $B = 2.2$, where $B_c = 1 + \sqrt{1 + 1/9} \approx 2.05$. Here fewer trials were used as the algorithm slows down for larger C, so the graphs are visibly less smooth. Again, the normal distribution is evident in the supercritical case.

9.2. **Single special row and column case.** As we mentioned in the introduction, this paper was motivated by Barvinok's question about the phase transition for a single special row and column (the case $\delta = 0$ in our notation). We are now confident in the conjecture that the phase transition occurs at $B = B_c$ based on our results and extensive simulations, see Figure 2 and [DLP19+]. In fact, we believe both Conjecture 3.1 and 3.2 extend to the single row and column case. Unfortunately, this case remains out of reach by the available tools.
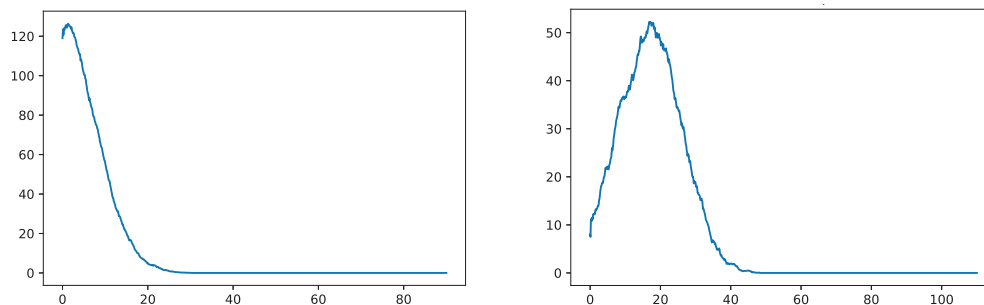
FIGURE 4. Summary of 10^3 simulations of x_{11}/C in a random uniform $n \times n$ contingency table $X = (x_{ij})$ with both margins $(BCn, Cn, \ldots, Cn)$, where $n = 50$ and $C = 9$. The left graph corresponds to a subcritical value $B = 1.8$, and the right graph to a supercritical value $B = 2.2$.

9.3. **Symmetric, binary, and high-dimensional tables.**

There are several variations on results in this paper one can consider. First, one can study *symmetric* contingency tables $X = X^T$ with the same row and column sums as in (2.1). These correspond to adjacency matrices of multigraphs with degree vectors $\mathbf{r}$. The results in this paper extend verbatim to this setting.

It would be even more interesting to understand the cases of general and symmetric binary (0-1) contingency tables (with zeros on the diagonal in the symmetric case), which correspond to simple bipartite and usual graphs. This is an extremely well studied family of problems, see [Wor18] for a recent thorough survey and [Bar10a, BH13] for the typical tables approach in this case. Note that since the entries are restricted to 0-1, the resulting distributions are Bernoulli but one needs to be careful in describing the phase transition. We believe our $\delta > 1/2$ results should be possible to modify in this case.

It would be also interesting to see if our results extend to higher dimensional contingency tables, see Remark 4.4. We reserve the judgement in this case; cf. [DP18] however.

9.4. **Criticality.**

The phase transition exhibited in the paper is somewhat different from known examples of phase transition in Statistical Mechanics, see [Yep18], Random Graph Theory, see [AS16, Bol01], or MCMC, see [Dia96]. It would be very interesting to understand exactly what happens at the critical values $B = B_c$ or $\delta = 1/2$. Unfortunately, these problems remain out of reach both theoretically and experimentally.

REFERENCES

[AS16] N. Alon and J. H. Spencer, *The probabilistic method* (Fourth ed.), John Wiley, Hoboken, NJ, 2016.

[Bar09] A. Barvinok, Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes, *Internat. Math. Res. Notices* **2009** (2009), 348–385.

[Bar10a] A. Barvinok, On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries, *Adv. Math.* **224** (2010), 316–339.

[Bar10b] A. Barvinok, What does a random contingency table look like?, *Comb. Probab. Comp.* **19** (2010), 517–539.

[Bar12] A. Barvinok, Matrices with prescribed row and column sums, *Lin. Alg. Appl.* **436** (2012), 820–844.

[BH10] A. Barvinok and J. A. Hartigan, Maximum entropy gaussian approximations for the number of integer points and volumes of polytopes, *Adv. Applied Math.* **45** (2010), 252–289.

[BH13] A. Barvinok and J. A. Hartigan, The number of graphs and a random graph with a given degree sequence, *Rand. Struct. Algor.* **42** (2013), 301–348.

[BLSY10] A. Barvinok, Z. Luria, A. Samorodnitsky and A. Yong, An approximation algorithm for counting contingency tables, *Rand. Struct. Algor.* **37** (2010), 25–66.

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

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

[BC78] 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.

[Bol01] B. Bollobás, *Random graphs* (Second ed.), Cambrie Univ. Press, Cambridge, UK, 2001.

[CM09] E. R. Canfield and B. D. McKay, The asymptotic volume of the Birkhoff polytope, *Online J. Anal. Comb.* **4** (2009), Paper 2, 4 pp.

[CM10] E. R. Canfield and B. D. McKay, Asymptotic enumeration of integer matrices with large equal row and column sums, *Combinatorica* **30** (2010), 655–680.

[CDS10] S. Chatterjee, P. Diaconis and A. Sly, Properties of uniform doubly stochastic matrices, preprint (2010), 18 pp.; `arXiv:1010.6136`.

[CV16] B. Cousins and S. Vempala, A practical volume algorithm, *Math. Program. Comput.* **8** (2016), 133–160.

[DK14] J. De Loera and E. D. Kim, Combinatorics and geometry of transportation polytopes: an update, in *Discrete geometry and algebraic combinatorics*, AMS, Providence, RI, 2014, 37–76.

[Dia96] P. Diaconis, The cutoff phenomenon in finite Markov chains, *Proc. Nat. Acad. Sci. U.S.A.* **93** (1996), no. 4, 1659–1664.

[DE85] P. Diaconis and B. Efron, Testing for independence in a two-way table: new interpretations of the chi-square statistic, *Ann. Stat.* **13** (1985), 845–913.

[DG95] P. Diaconis and A. Gangolli, Rectangular arrays with fixed margins, *Disc. Prob. Alg.* **72** (1995), 15–41.

[DS98] P. Diaconis and B. Sturmfels, Algebraic algorithms for sampling from conditional distributions, *Ann. Stat.* **26** (1998), 363–397.

[DP18] S. Dittmer and I. Pak, Random sampling and approximate counting of sparse contingency tables, submitted (2018).

[DLP19+] S. Dittmer, H. Lyu, and I. Pak, Contingency tables have empirical bias, in preparation (2019).

[Eve92] B. S. Everitt, *The analysis of contingency tables* (Second ed.), Chapman & Hall, London, 1992.

[FLL17] M. W. Fagerland, S. Lydersen and P. Laake, *Statistical analysis of contingency tables*, CRC Press, Boca Raton, FL, 2017.

[Goo63] I. J. Good, Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Annals Math. Stat.* **34** (1963), 911–934.

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

[GM08] C. Greenhill and B. D. McKay, Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums, *Adv. Appl. Math.* **41** (2008), 459–481.

[Kat14] M. Kateri, *Contingency table analysis*, Birkhäuser, New York, NY, 2014.

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

[Wor18] N. Wormald, Asymptotic enumeration of graphs with given degree sequence, in *Proc. ICM Rio de Janeiro*, Vol. 3, 2018, 3229–3248.

[Yep18] J. M. Yeomans, *Statistical Mechanics of Phase Transitions*, Oxford Univ. Press, 1992.

SAM DITTMER, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095, USA
Email address: `samuel.dittmer@math.ucla.edu`

HANBAEK LYU, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095, USA
Email address: `hlyu@math.ucla.edu`

IGOR PAK, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095, USA
Email address: `pak@math.ucla.edu`

APPENDIX

A. ASYMPTOTIC INDEPENDENCE BETWEEN ENTRIES

In this appendix, we show that any two entries in the bottom right block of $\mathcal{M}_{n,\delta}(B,C)$ are asymptotically independent in the sense that the correlation between any two entries of the uniform contingency table X decays polyonimally depending on the blocks that the entries belong.

Theorem A.1. *Fix* $\delta \in [0,1]$, $B,C \geq 0$, *and let* $X = (X_{ij})$ *be sampled from* $\mathcal{M}_{n,\delta}(B,C)$ *uniformly at random. Fix* $\varepsilon > 0$. *Then for every* $\varepsilon \geq 1$, *we have*

$$\sup_{a,b\in\mathbb{N}} \left| \mathbb{P}(X_{i_1,j_1} = a,\, X_{i_2,j_2} = b) - \mathbb{P}(X_{i_1,j_1} = a)\mathbb{P}(X_{i_2,j_2} = b) \right| \leq n^{-\eta(\delta)+\varepsilon}$$

for all sufficiently large $n \geq 1$, *where*

(A1)
$$\eta(\delta) = \begin{cases} 1/2 & \text{if all } (i_r, j_r)\text{'s are contained in the bottom right block,} \\ \delta - 1/2 & \text{if some } (i_r, j_r) \text{ is contained in the top left block,} \\ \delta/2 & \text{otherwise.} \end{cases}$$

The following lemma is a minor generalization of Lemma 4.2.

Lemma A.2 (Concentration of submatrices). *Let* $\mathcal{M}(\mathbf{r},\mathbf{c})$ *be the set of all* $m \times n$ *contingency tables of margins* $\mathbf{r}$ *and* $\mathbf{c}$. *Let* $X = (X_{ij})$ *be sampled from* $\mathcal{M}(\mathbf{r},\mathbf{c})$ *uniformly at random. Suppose* $\mathcal{B}_1,\ldots,\mathcal{B}_k$ *are blocks in* $\mathcal{M}(\mathbf{r},\mathbf{c})$ *with* $|\mathcal{B}_1| \leq \ldots \leq |\mathcal{B}_k|$. *Then there exists an absolute constant* $\gamma > 0$, *s.t. for each* $I = (I_1,\ldots,I_k) \in \mathcal{B}_1 \times \ldots \times \mathcal{B}_k$ *and* $t > 0$, *we have:*

$$d_{TV}\left((X_{I_\ell})_{\ell=1}^k, (Y_{I_\ell})_{\ell=1}^k \right) \leq t + N^{\gamma(m+n)} \exp\left(-\left\lfloor \frac{|\mathcal{B}_1|}{k} \right\rfloor \frac{t^2}{2} \right),$$

where $Z = (z_{ij})$ *is the typical table for* $\mathcal{M}(\mathbf{r},\mathbf{c})$, *and* $Y = (Y_J)$ *is the random matrix of independent entries with* $Y_{ij} \sim \text{Geom}(z_{ij})$, *and* $N = r_1 + \ldots + r_m = c_1 + \ldots + c_n$.

Proof. Let $\mathcal{M}(\mathbf{r},\mathbf{c})$ blocks $\mathcal{B}_1,\ldots,\mathcal{B}_k$ such that $|\mathcal{B}_1| \leq \ldots \leq |\mathcal{B}_k|$. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}(\mathbf{r},\mathbf{c})$, and let $Y = (Y_J)$ denote the random matrix of independent entries where $Y_{ij} \sim \text{Geom}(z_{ij})$. Observe that we can choose a subset $\mathcal{I} \subseteq \mathcal{B}_1 \times \ldots \times \mathcal{B}_k$ such that $|\mathcal{I}| \geq |\mathcal{B}_1|/k$ and every two elements of $\mathcal{I}$ have distinct coordinates. Fix measurable sets $A \subseteq \mathbb{R}_+^k$ and $\mathcal{A} \subseteq \mathbb{R}_+^{mn}$.

For a $m \times n$ matrix $W = (w_{ij})$ and $I \in \mathcal{B}_1 \times \ldots \times \mathcal{B}_k$, denote

$$W^I := (w_{I_\ell})_{\ell=1}^k \in \mathbb{R}^k, \qquad S(W) := \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \mathbf{1}_{\{W^I \in A\}}.$$

Note that by the exchangeability of the entries of X and Y in each block of $\mathcal{M}(\mathbf{r},\mathbf{c})$, random vectors X^I and Y^I have the same distribution for all I. In particular, we have:

$$\mathbb{P}(X^I \in A) = \mathbb{E}[S(X)].$$

Moreover, since Y_{ij} are independent and since every two elements of $\mathcal{I}$ have non-overlapping coordinates, it follows that Y^I are also independent.

Now the rest of the argument is identical to the proof of Lemma 4.2. $\square$

The following is a vector version of Theorem 6.1.

Theorem A.3. *Let* $X = (X_{ij})$ *be sampled from* $\mathcal{M}_{n,\delta}(B,C)$ *uniformly at random. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{M}_{n,\delta}(B,C)$ *and let* $Y = (Y_{ij})$ *denote the random matrix of independent entries with* $Y_{ij} \sim \text{Geom}(z_{ij})$. *Fix integers* $k \geq 1$ *and* $1 \leq i_r, j_r \leq n + \lfloor n^\delta \rfloor$, $1 \leq r \leq k$. *Then for every* $\varepsilon > 0$, *we have:*

$$d_{TV}\left((X_{i_r,j_r})_{r=1}^k, (Y_{i_r,j_r})_{r=1}^k \right) \leq n^{-\eta(\delta)+\varepsilon}$$

for all sufficiently large $n \geq 1$, where $\eta(\delta)$ is defined as in (A1).

Proof. Identical to that for Theorem 6.1 using Lemma A.2.   □

Proof of Theorem A.1. Let $Z = (z_{ij})$ be the typical table for $\mathscr{M}(\mathbf{r}, \mathbf{c})$, and let $Y = (Y_J)$ denote the random matrix of independent entries where $Y_{ij} \sim \text{Geom}(z_{ij})$. By the symmetry within blocks, in order to bound the maximum in the assertion, we only need to consider six cases corresponding to the combination of blocks that entries $X_{i,j}$ and $X_{k,\ell}$ belong.

Consider the case when both entries are contained in the top left blocks. The other cases can be shown by a similar argument. Note that for any $a, b \in \mathbb{N}$,

$$\left| \mathbb{P}(X_{n+1,n+1} = a, X_{n+2,n+2} = b) - \mathbb{P}(Y_{n+1,n+1} = a, Y_{n+2,n+2} = b) \right|$$
$$\leq d_{TV}((X_{n+1,n+1}, X_{n+2,n+2}), (Y_{n+1,n+1}, Y_{n+2,n+2})),$$
$$\left| \mathbb{P}(X_{n+1,n+1} = a) - \mathbb{P}(Y_{n+1,n+1} = a) \right| \leq d_{TV}(X_{n+1,n+1}, Y_{n+1,n+1}),$$
$$\left| \mathbb{P}(X_{n+2,n+2} = b) - \mathbb{P}(Y_{n+2,n+2} = b) \right| \leq d_{TV}(X_{n+2,n+2}, Y_{n+2,n+2}).$$

Since Y_{n+1} and Y_{n+2} are independent, triangle inequality gives

$$\sup_{a,b \in \mathbb{N}} \left| \mathbb{P}(X_{n+1,n+1} = a, X_{n+2,n+2} = b) - \mathbb{P}(X_{n+1,n+1} = a)\mathbb{P}(X_{n+2,n+2} = b) \right|$$
$$\leq d_{TV}((X_{n+1,n+1}, X_{n+2,n+2}), (Y_{n+1,n+1}, Y_{n+2,n+2})) + d_{TV}(X_{n+1,n+1}, Y_{n+1,n+1}) + d_{TV}(X_{n+2,n+2}, Y_{n+2,n+2}).$$

Then by Theorem 6.1, for any fixed $\varepsilon > 0$, each of the total variation distances above is at most $n^{-1/2+\varepsilon}$ for sufficiently large $n \geq 1$.   □