ALEA, Lat. Am. J. Probab. Math. Stat. 20, 1367–1383 (2023)
DOI: 10.30757/ALEA.v20-51

On Properties of Random Binary Contingency Tables with Non-Uniform Margin

Da Wu

Department of Mathematics, University of Pennsylvania, Philadelphia, PA, 19104-6395, USA
E-mail address: dawu@sas.upenn.edu
URL: https://sites.google.com/view/dawu/home

Abstract. In this paper, we considered the random binary contingency tables with non-uniform margin. More precisely, for parameters n, δ, B, C, let $X = (X_{ij})$ with $X_{ij} \in \{0, 1\}$ denote the uniform sample from the set of $([n^\delta] + n)$-dimensional binary matrices whose first $[n^\delta]$ rows and columns have margin $[BCn]$ and the remaining n columns and rows have margin $[Cn]$. Various asymptotic properties of X as $n \to \infty$ were obtained.

1. Introduction

1.1. *Overview.* In Statistics, contingency tables serve as a means to represent the interrelationship within extensive datasets. From a mathematical perspective, these tables can be described as a collection of matrices where both the row and column sums are predetermined. When the totals of rows and columns are contingent upon the dimensions of the matrix, grasping the asymptotic patterns exhibited by the contingency table as these dimensions expand becomes a formidable undertaking. Combinatorists are interested in deriving a precise asymptotic formula for the cardinality of the contingency tables. The study by Canfield and McKay (2010) employed the multi-variable Cauchy integral formula to address the scenario of uniform margins, where both the row and column sums are identical, often referred to as a *magic square* in Stanley (2012). Subsequently, Barvinok and Hartigan (2012) obtained a precise asymptotic formula for scenarios with non-uniform margins, utilizing the so-called *Maximum entropy principle.* Their fundamental approach involved representing the count of contingency tables through a probability density function evaluated at a specific fixed point, subsequently applying the local central limit theorem for approximation. Their formula remains valid under the condition that the entries of the *typical table*, which is a matrix representing the expected values of each entry as the limit is approached, remain within bounded limits as the limit is taken. The exact definition of the typical table will be postponed and provided in Definition 2.1. The case when some entries of the typical table blow up remain unsolved. For readers intrigued by the various combinatorial aspects of contingency tables, see the survey paper Diaconis and Gangolli (1995).

Received by the editors September 16th, 2022; accepted September 21st, 2023.
2010 *Mathematics Subject Classification.* 60F05, 60C05.
Key words and phrases. Random contingency table, Binary contingency table, Maximum Entropy Principle.

In Probability Theory, the collection of contingency tables serves as the finite probability space and we equip the uniform probability measure on this space. Notably, the sums of rows and columns are contingent upon the dimension of matrix. It is a natural inquiry to pose when dealing with a significantly large dimension: If we randomly sample a matrix from a uniform distribution, what will it resemble? In greater detail, what can we discern about the approaching marginal distribution of individual entries, as well as the joint distribution of sub-matrices? Additionally, can any insights be gained regarding the spectrum in this scenario? I. J. Good introduced the concept of the *Maximum Entropy Principle* in his work Good (1963), proposing its application in the analysis of random contingency tables. After half a century, A. Barvinok ultimately validated Good's initial insight, as detailed in Barvinok (2010b) and its related references. Despite significant advancements, the problem of determining the marginal distribution of individual entries remained unsolved across a wide range of scenarios. Chatterjee, Diaconis, and Sly achieved a solution for the class of *doubly stochastic matrices*, which are real matrices with equal row and column sums equating to 1, as documented in Chatterjee et al. (2010). The author, in his work Wu (2023b), addressed the discrete case with uniform margins. Dittmer, Lyu, and Pak, as discussed in Dittmer et al. (2020), delved into the non-uniform margin scenario, elucidating the precise phase transitions in terms of limiting behaviors. As applications, Lyu and Pak, in their work Lyu and Pak (2022), obtained sharp asymptotic estimates on the number of $n \times n$ contingency tables with non-uniform margins and showed that in the supercritical regime, the classical independence heuristic leads to a large over-counting.

In this paper, we focus on asymptotic properties of random binary contingency tables with non-uniform margins. It is worth noting that binary matrices with predetermined row and column sums hold significant relevance in various mathematical disciplines. For instance, in Combinatorics, they are closely tied to hypergraphs with fixed degrees of vertices and network flow analysis, as expounded upon in van Lint and Wilson (2001). Moreover, these binary matrices emerge naturally as structural constants in symmetric function theory, thus playing a pivotal role in the representation theory of symmetric and general linear groups, as elucidated in Macdonald (1998).

Through the maximization of *Shannon-Boltzmann entropy* for Bernoulli random variables while subject to first-order constraints, namely, fixed row and column sums, we have derived the limiting marginal distribution for the uniformly sampled binary contingency tables. Furthermore, we have demonstrated that the joint distribution of entries within each block tends toward a set of independent and identically distributed (i.i.d.) Bernoulli variables, corroborating the independence heuristic outlined in Good (1976); Good and Crook (1977). Finally, we have delved into the convergence rates of higher moments of entries within random binary contingency tables, establishing the validity of the strong law of large numbers for certain truncated rows.

1.2. *Basic setups.* Let $\mathbf{r} = (r_1, \ldots, r_m) \in \mathbb{N}^m$ and $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{N}^n$ be two positive integer vectors of length m and n respectively with the same total sum of entries N, i.e,

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

We call such vectors $\mathbf{r}$ and $\mathbf{c}$ margins. Let $\mathscr{M}^{\{0,1\}}(\mathbf{r}, \mathbf{c})$ denote the set of $m \times n$ binary contingency tables with row sums r_i and column sums c_j, i.e.,

$$\mathscr{M}^{\{0,1\}}(\mathbf{r}, \mathbf{c}) := \left\{ (d_{ij}) \in \{0,1\}^{mn} : \sum_{k=1}^{n} d_{ik} = r_i, \sum_{k=1}^{m} d_{kj} = c_j \text{ for all } 1 \le i \le m, 1 \le j \le n \right\}.$$

For $B, C > 0$ and $0 \le \delta \le 1$, the *Barvinok margins* $\widetilde{\mathbf{r}}$ and $\widetilde{\mathbf{c}}$ as in Dittmer et al. (2020); Lyu and Pak (2022) are defined as

$$\widetilde{\mathbf{r}} = \widetilde{\mathbf{c}} := (\underbrace{[BCn], \ldots, [BCn]}_{[n^\delta] \text{ entries}}, \underbrace{[Cn], \ldots, [Cn]}_{n \text{ entries}}) \in \mathbb{N}^{[n^\delta]+n}. \tag{1.1}$$

Furthermore, let

$$\mathscr{M}_{n,\delta}^{\{0,1\}}(B, C) := \mathscr{M}^{\{0,1\}}(\widetilde{\mathbf{r}}, \widetilde{\mathbf{c}}),$$

and let $X = (X_{ij})_{1 \le i,j \le n+[n^\delta]}$ be a uniform sample from $\mathscr{M}_{n,\delta}^{\{0,1\}}(B, C)$. We call the random sample X a *random binary contingency table*. The main objective of this paper is to study various asymptotic properties of X as $n \to \infty$.

First, the following trivial bounds on B and C need to hold so that the set $\mathscr{M}_{n,\delta}^{\{0,1\}}(B, C)$ is always non-empty as $n \to \infty$.

Lemma 1.1. *Suppose the set $\mathscr{M}_{n,\delta}^{\{0,1\}}(B, C)$ is non-empty, then we have*

$$\begin{cases} 0 < C \le 1 \\ 0 < B \le \frac{1}{C} \end{cases} \quad \text{if } 0 \le \delta < 1, \quad \text{and} \quad \begin{cases} 0 < C \le 2 \\ 0 < B \le \frac{2}{C} \end{cases} \quad \text{if } \delta = 1.$$

Proof: Since every entry of the matrix is restricted to $\{0, 1\}$, we have $BCn \le [n^\delta] + n$ and $Cn \le [n^\delta] + n$, which are equivalent to $BC \le 1 + \frac{[n^\delta]}{n}$ and $C \le 1 + \frac{[n^\delta]}{n}$. The results then follow from taking limits as $n \to \infty$. □

1.3. *Notations.*

(1) For two random variables X_1, X_2 taking values on $\mathbb{N}$, the *Total Variation Distance* metric is defined as
$$d_{TV}(X_1, X_2) := \sum_{k \ge 0} |\mathbb{P}(X_1 = k) - \mathbb{P}(X_2 = k)|.$$
(2) A random variable $X \sim \mathrm{Ber}(q)$ if $\mathbb{P}(X = 1) = q$ and $\mathbb{P}(X = 0) = 1 - q$.
(3) We use $f(n) = O(g(n))$ or $f \ll g$ to denote the estimate that there exist some $M > 0$ and real number x_0 such that
$$f(n) \le M \cdot g(n) \quad \text{for all } n \ge x_0.$$
(4) We use $f(n) = o(g(n))$ to denote the estimate that for any $\varepsilon > 0$, there exists a real number x_0 such that
$$f(n) \le \varepsilon \cdot g(n) \quad \text{for all } n \ge x_0.$$

1.4. *Main results.* In Dittmer et al. (2020), the authors established a sharp phase transition for typical table in terms of B for non-negative integer-valued contingency tables with the *Barvinok margins* (1.1). As a consequence, when incorporating additional constraints while extending the characteristics of typical tables to contingency tables, a similar phase transition emerges concerning the limiting behaviors of non-negative integer-valued random contingency tables.

Therefore, it is natural to inquire whether random *binary* contingency tables with Barvinok margins exhibit a comparable phase transition. While all the entries of the typical table in binary case cannot blow up as the super-critical case ($B > 1 + \sqrt{1 + 1/C}$) in Dittmer et al. (2020, Lemma 5.1), it is still not entirely clear whether limiting values of entries of the typical table do not exhibit any phase transition as as the ratio B between two margin values undergoes variation. In this work, we show that indeed there is *no* phase transition for entries of the typical table in binary case (Lemma 2.3). As a result, when incorporating some additional restrictions introduced by the estimate (3.1), there is *no* phase transition in terms of limiting behaviours of random binary contingency tables with Barvinok margins.

Theorem 1.2. *For $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$ with parameter n,δ,B,C, let $X=(X_{ij})$ be sampled uniformly at random from $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$. Fix $\varepsilon>0$ and we have the following:*

(1) *(Bottom right) When $0\le\delta<1$, $0<C\le 1$ and $0<B\le\frac{1}{C}$, we have*

$$d_{TV}\left(X_{n+1,n+1},\mathrm{Ber}(C)\right)=O\left(n^{\delta-1}+n^{-\frac{1}{2}+\varepsilon}\right).$$

(2) *(Top left) When $\frac{1}{2}<\delta<1$, $0<C\le 1$ and $0<B\le\frac{1}{C}$, we have*

$$d_{TV}\left(X_{11},\mathrm{Ber}\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right)=O\left(n^{\delta-1}+n^{\frac{1}{2}-\delta+\varepsilon}\right).$$

(3) *(Side blocks) When $0<\delta<1$, $0<C\le 1$ and $0<B\le\frac{1}{C}$, we have*

$$d_{TV}\left(X_{1,n+1},\mathrm{Ber}(BC)\right)=d_{TV}\left(X_{n+1,1},\mathrm{Ber}(BC)\right)=O\left(n^{\delta-1}+n^{-\frac{\delta}{2}+\varepsilon}\right).$$

It is trivial to see that

$$d_{TV}\left(\mathrm{Ber}(\lambda_1),\mathrm{Ber}(\lambda_2)\right)=2|\lambda_1-\lambda_2|=2\left|\mathbb{E}[\mathrm{Ber}^k(\lambda_1)]-\mathbb{E}[\mathrm{Ber}^k(\lambda_2)]\right|\quad\text{for all } k\ge 1,\tag{1.2}$$

which immediately implies the following corollary on convergence of higher moments of X:

Corollary 1.3. *For $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$ with parameter n,δ,B,C, let $X=(X_{ij})$ be sampled uniformly at random from $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$. Fix $\varepsilon>0$ and for all $k\ge 1$, we have the following:*

(1) *(Bottom right) When $0\le\delta<1$, $0<C\le 1$ and $0<B\le\frac{1}{C}$, we have*

$$\left|\mathbb{E}\left[X_{n+1,n+1}^k\right]-C\right|=O\left(n^{\delta-1}+n^{-\frac{1}{2}+\varepsilon}\right).$$

(2) *(Top left) When $\frac{1}{2}<\delta<1$, $0<C\le 1$ and $0<B\le\frac{1}{C}$, we have*

$$\left|\mathbb{E}\left[X_{11}^k\right]-\frac{B^2(1-C)}{B^2-2B+1/C}\right|=O\left(n^{\delta-1}+n^{\frac{1}{2}-\delta+\varepsilon}\right).$$

(3) *(Side blocks) When $0<\delta<1$, $0<C\le 1$ and $0<B\le\frac{1}{C}$, we have*

$$\left|\mathbb{E}[X_{1,n+1}^k]-BC\right|=O\left(n^{\delta-1}+n^{-\frac{\delta}{2}+\varepsilon}\right).$$

Our next result deals with the joint distribution of entries within each block. For $k=k(n)$ random variables $R_1,\dots,R_k$, let the vector $(R_1,\dots,R_k)$ denote their joint distribution. Let $V_k(\gamma)$ denote the joint distribution of k i.i.d. $\mathrm{Ber}(\gamma)$ variables.

Theorem 1.4. *For $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$ with parameter n,δ,B,C, let $X=(X_{ij})$ be sampled uniformly at random from $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$. Then we have the following:*

(1) *(Bottom right) When $0<C\le 1$, $0<B\le\frac{1}{C}$, and*

$$k=k(n)=\begin{cases}o\left(\frac{n}{\log n}\right) & \text{if } \delta=0,\\ o(n^{1-\delta}), & \text{if } 0<\delta<1,\end{cases}$$

we have

$$d_{TV}\left((X_{[n^\delta]+1,[n^\delta]+1},X_{[n^\delta]+1,[n^\delta]+2},\dots,X_{[n^\delta]+1,[n^\delta]+k}),V_k(C)\right)\to 0.$$

(2) *(Top left) When* $0 < C \le 1$, $0 < B \le \frac{1}{C}$, *and*

$$k = k(n) = \begin{cases} o\left(\frac{n^{2\delta-1}}{\log n}\right) & \text{if } \frac{1}{2} < \delta \le \frac{2}{3}, \\ o(n^{1-\delta}), & \text{if } \frac{2}{3} < \delta < 1, \end{cases}$$

we have

$$d_{TV}\left((X_{11}, \ldots, X_{1k}), V_k\left(\frac{B^2(1-C)}{B^2 - 2B + 1/C}\right)\right) \to 0.$$

(3) *(Side blocks)When* $0 < C \le 1$, $0 < B \le \frac{1}{C}$, *and*

$$k = k(n) = \begin{cases} o\left(\frac{n^{\delta}}{\log n}\right) & \text{if } 0 < \delta \le \frac{1}{2}, \\ o(n^{1-\delta}), & \text{if } \frac{1}{2} < \delta < 1, \end{cases}$$

we have

$$d_{TV}\left((X_{1,[n^\delta]+1}, \ldots, X_{1,[n^\delta]+k}), V_k(BC)\right) \to 0.$$

Therefore, within each block, the joint distribution of any $k = k(n)$ entries are asymptotically independen; In particular, it is true for any fixed number of entries.

Our final results deal with the Strong Law of Large Numbers (SLLN) for certain truncated rows. Let

$$S^{\mathbf{S}}_{n,\delta}(B,C) := \sum_{k=1}^{n} X_{1,k+[n^\delta]} \qquad \text{and} \qquad S^{\mathbf{BR}}_{n,\delta}(B,C) := \sum_{k=1}^{n} X_{n+1,k+[n^\delta]},$$

and we have the following two SLLN theorems:

Theorem 1.5. *Let* $0 < \delta \le \frac{1}{2}$, $0 < C \le 1$ *and* $0 < B \le \frac{1}{C}$ *and let* $X = (X_{ij})$ *be the uniform sample from* $\mathscr{M}^{\{0,1\}}_{n,\delta}(B,C)$. *Then we have*

$$\frac{1}{n} S^{\boldsymbol{S}}_{n,\delta}(B,C) \to BC \qquad \text{almost surely.}$$

Theorem 1.6. *Let* $0 \le \delta < 1$, $0 < C \le 1$ *and* $0 < B \le \frac{1}{C}$ *and let* $X = (X_{ij})$ *be the uniform sample from* $\mathscr{M}^{\{0,1\}}_{n,\delta}(B,C)$. *Then we have*

$$\frac{1}{n} S^{\boldsymbol{BR}}_{n,\delta}(B,C) \to C \qquad \text{almost surely.}$$

Remark 1.7. Notice that the expected values of $S^{\mathbf{S}}_{n,\delta}(B,C)$ converge to BCn, which is already the entire row sum. Hence, we do not expect the Central Limit Theorem (CLT) holds for $S^{\mathbf{S}}_{n,\delta}(B,C)$ since there is no room for the row sum to fluctuate. Likewise for $S^{\mathbf{BR}}_{n,\delta}(B,C)$. Similar reasoning was also mentioned in Dittmer et al. (2020, Conjecture 3.1), where in the sub-critical case ($B < 1 + \sqrt{1 + 1/C}$), the expected row rum is already equal to the full row margin and therefore the CLT is not expected to hold. On the contrary, in the super-critical case ($B > 1 + \sqrt{1 + 1/C}$), the expected row sum is strictly less than the whole row margin and we do expect the CLT to hold after suitable renormalizations.

1.5. *Open problems and future research.*

(1) In Nguyen (2014), the author showed that for uniformly doubly stochastic matrices, the empirical eigenvalue distribution converges to the circular law. Nevertheless, when dealing with non-uniform margins, the behavior of the spectrum remains elusive, and we lack even a speculative limit hypothesis.

(2) A straightforward extension of our findings involves examining scenarios where the elements of matrix X can assume values within $\{0, 1, \ldots, k\}$. The Maximum Entropy Principle suggests that the limiting distribution of X is likely to be a truncated Geometric distribution. Therefore, as $k \to \infty$, it should recover the results in Dittmer et al. (2020).

(3) Moreover, exploring the monotonic behavior of entropy becomes even more intriguing, as it appears that entropy consistently rises with each incremental increase in both k and n. This curiosity finds its inspiration in the well-known *Shannon Monotonicity Conjecture* related to the classical CLT, a conjecture that was substantiated in Artstein et al. (2004) through the application of Fisher information. It is feasible to explore an extension of this conjecture into higher dimensions using random contingency tables. This extension offers a broader scope as it not only involves an increase in dimensions but also entails a change in the nature of constraints. For instance, in higher dimensions, second-order constraints or variance constraints correspond to the use of uniformly distributed (also known as *Haar distributed*) orthogonal or unitary matrices. This category of problems has been extensively studied, and it is established that the marginal distribution of uniformly distributed orthogonal or unitary matrices converges to the standard normal distribution when appropriately rescaled to have a mean of 1. For further insights, readers can refer to works such as Diaconis and Shahshahani (1994); Borel (1914) and their associated references.

2. Asymptotic analysis of the typical table

A. Barvinok introduced the notion of *typical table* in order to answer the question: *What does a random contingency tables look like?* As the dimension of matrix increases, it becomes apparent that the random contingency table exhibits a degree of similarity, in a specific context, to the typical table. For further context and precise details, please refer to references Barvinok (2010b,a, 2012); Barvinok and Hartigan (2010). Here, we'll simply revisit the constructions by Barvinok and provide some additional observations.

Fix margins $\mathbf{r} \in \mathbb{N}^m$ and $\mathbf{c} \in \mathbb{N}^n$, the *binary transportation polytope* is defined to be

$$\mathscr{P}^{\{0,1\}}(\mathbf{r}, \mathbf{c}) := \left\{ (x_{ij}) \in [0,1]^{mn} : \sum_{k=1}^{m} x_{ik} = r_i, \sum_{k=1}^{n} x_{kj} = c_j, \forall 1 \le i \le m, 1 \le j \le n \right\}.$$

Definition 2.1 (Typical Table)**.** For all $X = (x_{ij}) \in (0,1)^{mn}$, let

$$g(X) := \sum_{i,j} x_{ij} \ln \frac{1}{x_{ij}} + (1 - x_{ij}) \ln \frac{1}{1 - x_{ij}}.$$

For fixed margin $\mathbf{r}$ and $\mathbf{c}$, we define the *typical table* $Z = (z_{ij})$ to be the unique maximizer of g in the interior of $\mathscr{P}^{\{0,1\}}(\mathbf{r}, \mathbf{c})$.

Remark 2.2. Note that

(1) For fixed i, j, the quantity $x_{ij} \ln \frac{1}{x_{ij}} + (1 - x_{ij}) \ln \frac{1}{1 - x_{ij}}$ is the *Shannon-Boltzmann entropy* Shannon (1948) of the $\mathrm{Ber}(x_{ij})$.

(2) Since g is strictly concave in the interior of $\mathscr{P}^{\{0,1\}}(\mathbf{r}, \mathbf{c})$, which is compact, the function g attains the unique maximum in that region. Therefore Definition 2.1 is well-defined.

(3) For fixed i, j and notice that $\frac{\partial}{\partial x_{ij}} g(X) = \ln\left(\frac{1 - x_{ij}}{x_{ij}}\right)$. Therefore, for typical table $Z = (z_{ij})$, we have the following *Lagrange multiplier conditions*:

$$\ln\left(\frac{1 - z_{ij}}{z_{ij}}\right) = \lambda_i + \mu_j, \quad \text{for } 1 \le i \le m \text{ and } 1 \le j \le n.$$

Next, we delve into the asymptotic behavior of the elements within the typical table denoted as $Z = (z_{ij})$. Our approach in this section closely aligns with the reasoning presented in Dittmer et al. (2020, Lemma 5.1, Proposition 5.2). The primary distinction lies in our optimization method, which relies on the entropy of Bernoulli random variables instead of Geometric random variables.

By symmetry and the theory of Lagrange multiplier, there exist some α, β (possibly depend on all the parameters) such that

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

Let $P = e^\alpha$ and $Q = e^\beta$ and we have

$$z_{ij} = \begin{cases} \frac{1}{P^2+1} & 1 \le i,j \le [n^\delta], \\ \frac{1}{Q^2+1} & [n^\delta] < i,j \le [n^\delta]+n, \\ \frac{1}{PQ+1} & \text{otherwise.} \end{cases} \tag{2.1}$$

We also have the following two marginal conditions for $Z = (z_{ij})$:

$$\begin{cases} ([n^\delta]/n)z_{11} + z_{1,n+1} = BC, \\ ([n^\delta]/n)z_{1,n+1} + z_{n+1,n+1} = C. \end{cases} \tag{2.2}$$

From (2.2), we have

$$\begin{cases} z_{n+1,n+1} \le C, \\ z_{1,n+1} \le BC, \end{cases} \quad \text{and} \quad \begin{cases} z_{n+1,n+1} = C + O(n^{\delta-1}), \\ z_{1,n+1} = BC + O(n^{\delta-1}), \end{cases} \tag{2.3}$$

since $0 < z_{ij} < 1$ for all $1 \le i,j \le n + [n^\delta]$.

Lemma 2.3. *Let $0 < C \le 1$ and $0 < B \le \frac{1}{C}$. Let $Z = (z_{ij})$ be the typical table for $\mathscr{M}_{n,\delta}^{\{0,1\}}(B,C)$ with $0 \le \delta < 1$, then we have*

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

Proof: Since z_{11} is uniformly bounded in n, we have

$$|z_{1,n+1} - BC| \le n^{\delta-1} z_{11} = O(n^{\delta-1}), \tag{2.4}$$

which implies that $\lim_{n\to\infty} z_{1,n+1} = BC$. Let $P = P(n), Q = Q(n)$ be as in (2.1), then

$$\lim_{n\to\infty} z_{1,n+1} = \lim_{n\to\infty} \frac{1}{PQ+1} = BC \qquad \text{and} \qquad \lim_{n\to\infty} z_{n+1,n+1} = \lim_{n\to\infty} \frac{1}{Q^2+1} = C,$$

which are equivalent to

$$Q \to q^* := \sqrt{\frac{1}{C} - 1} \qquad \text{and} \qquad PQ \to \frac{1}{BC} - 1.$$

Consequently, we have

$$P \to p^* := \left(\frac{1}{BC} - 1\right) \Big/ \sqrt{\frac{1}{C} - 1}$$

and

$$z_{11} = \frac{1}{P^2+1} \to \frac{1}{(p^*)^2+1} = \frac{B^2(1-C)}{B^2-2B+1/C} \le B^2 C.$$

Next, we want to obtain the convergence rate for z_{11}. Let $h(x) = \frac{1}{x^2+1}$ and $h'(x) = \frac{-2x}{(x^2+1)^2}$. Since $|h'(x)| \le \frac{3\sqrt{3}}{8}$ for all $x \in \mathbb{R}$, by Mean Value Theorem (see e.g. Lang (1998)), we have

$$\begin{aligned} |z_{11} - h(p^*)| = |h(P) - h(p^*)| &\le \sup_{x\in\mathbb{R}} |h'(x)| \cdot |P - p^*| \\ &= \frac{3\sqrt{3}}{8} \cdot |P - p^*|. \end{aligned}$$

Next, by triangle inequality,

$$|P - p^*| \le \left| P - \frac{1/BC - 1}{Q} \right| + \left(\frac{1}{BC} - 1 \right) \left| \frac{1}{Q} - \frac{1}{q^*} \right|. \tag{2.5}$$

Next, since $z_{n+1,n+1} = h(Q), C = h(q^*)$, by Mean Value Theorem,

$$BCn^{\delta-1} \ge |z_{n+1,n+1} - C| = |h(Q) - h(q^*)| \ge \frac{3\sqrt{3}}{8} \cdot |Q - q^*|$$

for sufficiently large n. Hence, $|Q - q^*| = O(n^{\delta-1})$. Since $Q \to q^*$, the second term in (2.5) is of order $O(n^{\delta-1})$. For the first term in (2.5), we have

$$\begin{aligned} \left| P - \frac{1/BC - 1}{Q} \right| &= \frac{(PQ+1)/BC}{Q} \cdot \left| \frac{1}{PQ+1} - BC \right| \\ &= \frac{(PQ+1)/BC}{Q} \cdot |z_{1,n+1} - BC| \\ &= O(n^{\delta-1}) \end{aligned} \tag{2.6}$$

since both P and Q converge as $n \to \infty$ and by (2.4), we have $|z_{1,n+1} - BC| = O(n^{\delta-1})$. Thus $|P - p^*| = O(n^{\delta-1})$ and this completes the proof. $\square$

As an application of the Lemma 2.3, the author in Wu (2023a) obtained a sharp asymptotic formula for the number of binary contingency tables with Bavinok margins (1.1), following closely from the work of Lyu and Pak (2022). Compared with Lyu and Pak (2022, Theorem 2.2), where there exist a sharp phase transition in terms of the asymptotics on the number of non-negative integer valued contingency tables with Barvinok margins, there is *no* phase transition in the binary case Wu (2023a, Theorem 1.2).

Furthermore, it is quite interesting to see that, in the binary case, the independence heuristic, as defined in Wu (2023a, (2)), is proven to *overestimate* by a large factor, whereas in the non-negative integer valued case, the independence heuristic, as defined in Lyu and Pak (2022, (1.3)), leads to a large *under-counting*.

3. Estimations on total variation distance and Proof of Theorems 1.2 and 1.4

In this section, we use concentration inequality to prove Theorems 1.2 and 1.4. The proof is verbatim to that of Dittmer et al. (2020, Theorem 2.1) and Chatterjee et al. (2010, Theorem 1). First, we recall the following theorem in Barvinok (2012).

Theorem 3.1 (Barvinok (2012)). *Fix row margin* $\mathbf{r} = (r_1, \dots, r_m)$ *and column margin* $\mathbf{c} = (c_1, \dots, c_n)$. *Let* $Z = (z_{ij})_{1\le i\le m, 1\le j\le n}$ *be the typical table for* $\mathscr{M}^{\{0,1\}}(\mathbf{r}, \mathbf{c})$. *Let* $Y = (y_{ij})_{1\le i\le m, 1\le j\le n}$ *be an matrix with independent Bernoulli random variables with* $y_{ij} \sim Ber(z_{ij})$. *Then we have the following conclusions:*

(1) *There exists an absolute constant* γ *such that*

$$(mn)^{-\gamma(m+n)} e^{g(Z)} \le \left| \mathscr{M}^{\{0,1\}}(\mathbf{r}, \mathbf{c}) \right| \le e^{g(Z)}.$$

(2) *Conditioned on being in* $\mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})$*, the matrix* Y *is uniform on* $\mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})$*. In other words, the probability mass function of* Y *is constant on the set* $\mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})$*. More precisely, for any* $D \in \mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})$*, we have* $\mathbb{P}(Y = D) = e^{-g(Z)}$*.*

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

$$\mathbb{P}\left(Y \in \mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})\right) = e^{-g(Z)} \cdot \left|\mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})\right| \geq (mn)^{-\gamma(m+n)}.$$

Remark 3.2. By Theorem 3.1, for any fixed measurable set $\mathscr{A} \subseteq [0,1]^{mn}$, we have

$$\begin{aligned}\mathbb{P}(Y \in \mathscr{A}) &\geq \mathbb{P}\left(Y \in \mathscr{A} | Y \in \mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})\right) \cdot \mathbb{P}\left(Y \in \mathscr{M}^{\{0,1\}}(\mathbf{r},\mathbf{c})\right) \\ &\geq \mathbb{P}(X \in \mathscr{A}) \cdot (mn)^{-\gamma(m+n)}.\end{aligned} \tag{3.1}$$

Subsequently, our goal is to derive an approximation for the total variation distance between the elements of the uniformly sampled matrix X and the maximum entropy matrix Y. In this endeavor, we employ the identical large deviation estimation method as introduced in Dittmer et al. (2020); Chatterjee et al. (2010).

Lemma 3.3. *Let* $X = (X_{ij})$ *be uniformly distributed on* $\mathscr{M}^{\{0,1\}}_{n,\delta}(B,C)$ *and* $Z = (z_{ij})$ *be the typical table associated to* $\mathscr{M}^{\{0,1\}}_{n,\delta}(B,C)$*. Let* $Y = (Y_{ij})$ *be the matrix of independent Bernoulli random variables with mean* z_{ij}*, i.e.* $Y_{ij} \sim \mathrm{Ber}(z_{ij})$*. Then, for any fixed* $\varepsilon > 0$*, we have*

$$\begin{cases} d_{TV}(X_{11}, Y_{11}) = O\left(n^{\frac{1}{2}-\delta+\varepsilon}\right), \\ d_{TV}(X_{1,n+1}, Y_{1,n+1}) = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right), \\ d_{TV}(X_{n+1,1}, Y_{n+1,1}) = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right), \\ d_{TV}(X_{n+1,n+1}, Y_{n+1,n+1}) = O\left(n^{-\frac{1}{2}+\varepsilon}\right). \end{cases}$$

Proof: By exchangeability of entries in the top left block and the Azuma-Hoeffding inequality, for any fixed measurable set $\mathcal{A} \subseteq [0,\infty)$, we have

$$\mathbb{P}\left(\left|\frac{1}{[n^\delta]^2}\sum_{1\leq i\leq [n^\delta]}\sum_{1\leq j\leq [n^\delta]} \mathbb{1}_{\{Y_{ij}\in\mathcal{A}\}} - \mathbb{P}\left(Y_{11}\in\mathcal{A}\right)\right| > t\right) \leq \exp\left(-2t^2[n^\delta]^2\right).$$

Moreover, by (3.1),

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{[n^\delta]^2}\sum_{1\leq i\leq [n^\delta]}\sum_{1\leq j\leq [n^\delta]} \mathbb{1}_{\{X_{ij}\in\mathcal{A}\}} - \mathbb{P}\left(Y_{11}\in\mathcal{A}\right)\right| > t\right) \\ &\leq \left(n+[n^\delta]\right)^{\gamma'(n+[n^\delta])} \cdot \exp\left(-2t^2\left([n^\delta]\right)^2\right)\end{aligned}$$

for some absolute constant $\gamma' > 0$. Therefore, we have

$$
\begin{aligned}
&|\mathbb{P}(X_{11} \in \mathcal{A}) - \mathbb{P}(Y_{11} \in \mathcal{A})| \\
&= \left| \mathbb{E}\left[\frac{1}{[n^\delta]^2} \sum_{1 \le i,j \le [n^\delta]} \mathbb{1}_{\{X_{ij} \in \mathcal{A}\}} \right] - \mathbb{P}(Y_{11} \in \mathcal{A}) \right| \\
&\le \mathbb{E}\left[\left| \frac{1}{[n^\delta]^2} \sum_{1 \le i,j \le [n^\delta]} \mathbb{1}_{\{X_{ij} \in \mathcal{A}\}} - \mathbb{P}(Y_{11} \in \mathcal{A}) \right| \right] \\
&\le t\mathbb{P}\left(\left| \frac{1}{[n^\delta]^2} \sum_{1 \le i,j \le [n^\delta]} \mathbb{1}_{\{X_{ij} \in \mathcal{A}\}} - \mathbb{P}(Y_{11} \in \mathcal{A}) \right| \le t \right) \\
&+ 2\mathbb{P}\left(\left| \frac{1}{[n^\delta]^2} \sum_{1 \le i,j \le [n^\delta]} \mathbb{1}_{\{X_{ij} \in \mathcal{A}\}} - \mathbb{P}\left(Y_{11} \in \mathcal{A}\right) \right| > t \right) \\
&\le t + 2\left(n + [n^\delta]\right)^{\gamma'(n+[n^\delta])} \cdot \exp\left(-2t^2 [n^\delta]^2\right).
\end{aligned}
$$

Fix $\varepsilon > 0$ and let $t = n^{\frac{1}{2}-\delta+\varepsilon}$. Then

$$
|\mathbb{P}(X_{11} \in \mathcal{A}) - \mathbb{P}(Y_{11} \in \mathcal{A})| = O\left(n^{\frac{1}{2}-\delta+\varepsilon}\right).
$$

By the exactly same argument,

$$
|\mathbb{P}(X_{1,n+1} \in \mathcal{A}) - \mathbb{P}(Y_{1,n+1} \in \mathcal{A})| \le t + 2\left(n + [n^\delta]\right)^{\gamma''(n+[n^\delta])} \cdot \exp\left(-2t^2 \cdot [n^\delta] \cdot n\right).
$$

Let $t = n^{-\frac{\delta}{2}+\varepsilon}$, then

$$
|\mathbb{P}(X_{1,n+1} \in \mathcal{A}) - \mathbb{P}(Y_{1,n+1} \in \mathcal{A})| = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right).
$$

Finally,

$$
|\mathbb{P}(X_{n+1,n+1} \in \mathcal{A}) - \mathbb{P}(Y_{n+1,n+1} \in \mathcal{A})| \le t + 2\left(n + [n^\delta]\right)^{\gamma'''(n+[n^\delta])} \cdot \exp\left(-2t^2 \cdot n^2\right).
$$

Let $t = n^{-\frac{1}{2}+\varepsilon}$, then

$$
|\mathbb{P}(X_{n+1,n+1} \in \mathcal{A}) - \mathbb{P}(Y_{n+1,n+1} \in \mathcal{A})| = O\left(n^{-\frac{1}{2}+\varepsilon}\right).
$$

Therefore, we have

$$
\begin{cases}
d_{TV}(X_{11}, Y_{11}) = O\left(n^{\frac{1}{2}-\delta+\varepsilon}\right), \\
d_{TV}(X_{1,n+1}, Y_{1,n+1}) = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right), \\
d_{TV}(X_{n+1,1}, Y_{n+1,1}) = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right), \\
d_{TV}(X_{n+1,n+1}, Y_{n+1,n+1}) = O\left(n^{-\frac{1}{2}+\varepsilon}\right).
\end{cases}
$$

This completes the proof. $\square$

Proof of Theorem 1.2: By (1.2), Lemma 2.3 and (2.3), we have

$$\begin{cases} d_{TV}\left(\mathrm{Ber}(z_{n+1,n+1}), \mathrm{Ber}(C)\right) = 2|z_{n+1,n+1} - C| = O(n^{\delta-1}), \\ d_{TV}\left(\mathrm{Ber}(z_{1,n+1}), \mathrm{Ber}(BC)\right) = 2|z_{1,n+1} - BC| = O(n^{\delta-1}), \\ d_{TV}\left(\mathrm{Ber}(z_{n+1,1}), \mathrm{Ber}(BC)\right) = 2|z_{n+1,1} - BC| = O(n^{\delta-1}), \\ d_{TV}\left(\mathrm{Ber}(z_{11}), \mathrm{Ber}\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) = 2\left|z_{11} - \frac{B^2(1-C)}{B^2-2B+1/C}\right| = O(n^{\delta-1}). \end{cases} \tag{3.2}$$

By Lemma 3.3,

$$\begin{cases} d_{TV}(X_{11}, Y_{11}) = O\left(n^{\frac{1}{2}-\delta+\varepsilon}\right), \\ d_{TV}(X_{1,n+1}, Y_{1,n+1}) = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right), \\ d_{TV}(X_{n+1,1}, Y_{n+1,1}) = O\left(n^{-\frac{\delta}{2}+\varepsilon}\right), \\ d_{TV}(X_{n+1,n+1}, Y_{n+1,n+1}) = O\left(n^{-\frac{1}{2}+\varepsilon}\right). \end{cases} \tag{3.3}$$

Therefore, by triangle inequality, we have

$$\begin{aligned} d_{TV}(X_{n+1,n+1}, \mathrm{Ber}(C)) &\le d_{TV}\left(X_{n+1,n+1}, \mathrm{Ber}\left(z_{n+1,n+1}\right)\right) + d_{TV}\left(\mathrm{Ber}(z_{n+1,n+1}), \mathrm{Ber}(C)\right) \\ &= O\left(n^{\delta-1} + n^{-\frac{1}{2}+\varepsilon}\right), \\ d_{TV}(X_{1,n+1}, \mathrm{Ber}(BC)) &\le d_{TV}\left(X_{1,n+1}, \mathrm{Ber}\left(z_{1,n+1}\right)\right) + d_{TV}\left(\mathrm{Ber}(z_{1,n+1}), \mathrm{Ber}(BC)\right) \\ &= O\left(n^{\delta-1} + n^{-\frac{\delta}{2}+\varepsilon}\right), \\ d_{TV}(X_{n+1,1}, \mathrm{Ber}(BC)) &\le d_{TV}\left(X_{n+1,1}, \mathrm{Ber}\left(z_{n+1,1}\right)\right) + d_{TV}\left(\mathrm{Ber}(z_{n+1,1}), \mathrm{Ber}(BC)\right) \\ &= O\left(n^{\delta-1} + n^{-\frac{\delta}{2}+\varepsilon}\right), \end{aligned}$$

and

$$\begin{aligned} &d_{TV}\left(X_{11}, \mathrm{Ber}\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) \\ &\le d_{TV}\left(X_{11}, \mathrm{Ber}\left(z_{11}\right)\right) + d_{TV}\left(\mathrm{Ber}(z_{11}), \mathrm{Ber}\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) \\ &= O\left(n^{\delta-1} + n^{\frac{1}{2}-\delta+\varepsilon}\right). \end{aligned}$$

This completes the proof. □

Proof of Theorem 1.4: We first prove (1). For $k = k(n)$, let $\mathcal{A} \subseteq \mathbb{R}^k$ be a fixed measurable subset and let

$$\mathscr{X}^{(\ell)} = \{(i_1^{(\ell)}, j_1^{(\ell)}), \ldots, (i_k^{(\ell)}, j_k^{(\ell)}) : [n^\delta] + 1 \le i_r^{(\ell)}, j_r^{(\ell)} \le [n^\delta] + n\}$$

be a k-subset of indices of the *bottom right* block with $\mathscr{X}^{(\ell)} \cap \mathscr{X}^{(\ell')} = \emptyset$ if $\ell \neq \ell'$ and $1 \le \ell \le [n^2/k]$. In other words, we divide the bottom right block into $[n^2/k]$ disjoint subsets, each with size k.

Let $X^{(\ell)} = (X_{(i_1^{(\ell)}, j_1^{(\ell)})}, \ldots, X_{(i_k^{(\ell)}, j_k^{(\ell)})})$ be the random vector of k entries, indexed by $\mathscr{X}^{(\ell)}$, in the bottom right block of X. By symmetry, $X^{(\ell)}$ has the same distribution with $X^{(\ell')}$ for all $1 \le \ell, \ell' \le [n^2/k]$. Similarly, let $Y^{(\ell)} = (Y_{(i_1^{(\ell)}, j_1^{(\ell)})}, \ldots, Y_{(i_k^{(\ell)}, j_k^{(\ell)})})$, where $Y = (Y_{ij})_{1 \le i,j \le n+[n^\delta]}$ is the matrix of independent Bernoulli random variables with $Y_{ij} \sim \mathrm{Ber}(z_{ij})$. By the Azuma-Hoeffding inequality, we have

$$\mathbb{P}\left(\left|\frac{1}{[n^2/k]} \sum_{\ell=1}^{[n^2/k]} \mathbb{1}_{\{X^{(\ell)} \in \mathcal{A}\}} - \mathbb{P}\left(Y^{(1)} \in \mathcal{A}\right)\right| > \frac{1}{2}\varepsilon\right) \le c' \frac{\exp(n \log n)}{\exp\left(-\frac{\varepsilon^2}{8} \cdot [n^{2\delta}/k]\right)}$$

for sufficiently large n. Hence, when $k = o\left(\frac{n}{\log n}\right)$, we have

$$\mathbb{P}\left(\left|\frac{1}{[n^2/k]}\sum_{\ell=1}^{[n^2/k]}\mathbb{1}_{\{X^{(\ell)}\in\mathcal{A}\}} - \mathbb{P}\left(Y^{(1)}\in\mathcal{A}\right)\right| > \frac{1}{2}\varepsilon\right) = o(1).$$

Since all $X^{(\ell)}$'s have the same distribution for $1 \le \ell \le [n^2/k]$, we have

$$\left|\mathbb{P}(X^{(1)}\in\mathcal{A}) - \mathbb{P}(Y^{(1)}\in\mathcal{A})\right| \le \frac{1}{2}\varepsilon + o(1).$$

Since $Y^{(1)}$ is a random vector of k independent Bernoulli random variables, we have

$$\begin{aligned} d_{TV}\left(Y^{(1)}, V_k(C)\right) &\le k \cdot d_{TV}\left(\mathrm{Ber}(z_{n+1,n+1}), \mathrm{Ber}(C)\right) \\ &= 2k|z_{n+1,n+1} - C|. \end{aligned}$$

By (3.2), we have $|z_{n+1,n+1}-C| = O(n^{\delta-1})$. Hence, when $k = o\left(n^{1-\delta}\right)$, we have $d_{TV}(Y^{(1)}, V_k(C)) = o(1)$. Notice that

$$\begin{cases} 1 \ll \frac{n}{\log n} \ll n^{1-\delta} & \text{if } \delta = 0, \\ 1 \ll n^{1-\delta} \ll \frac{n}{\log n} & \text{if } 0<\delta<1. \end{cases}$$

Hence, by triangle inequality, when

$$k = k(n) = \begin{cases} o\left(\frac{n}{\log n}\right) & \text{if } \delta = 0, \\ o(n^{1-\delta}) & \text{if } 0<\delta<1, \end{cases}$$

we have

$$d_{TV}\left((X_{[n^\delta]+1,[n^\delta]+1}, X_{[n^\delta]+1,[n^\delta]+2}, \ldots, X_{[n^\delta]+1,[n^\delta]+k}), V_k(C)\right) \to 0.$$

This proves (1).

Next, we prove (2). For $k = k(n)$, let $\mathcal{A} \subseteq \mathbb{R}^k$ be a fixed measurable subset. This time, let

$$\mathscr{X}^{(\ell)} = \{(i_1^{(\ell)}, j_1^{(\ell)}), \ldots, (i_k^{(\ell)}, j_k^{(\ell)}) : 1 \le i_r^{(\ell)}, j_r^{(\ell)} \le [n^\delta]\}$$

be a k-subset of indices of *top left block* with $\mathscr{X}^{(\ell)} \cap \mathscr{X}^{(\ell')} = \emptyset$ if $\ell \ne \ell'$ and $1 \le \ell \le [n^{2\delta}/k]$. In other words, we divide the top left block into $[n^{2\delta}/k]$ disjoint subsets, each with size k.

Let $X^{(\ell)} = (X_{(i_1^{(\ell)}, j_1^{(\ell)})}, \ldots, X_{(i_k^{(\ell)}, j_k^{(\ell)})})$ be the random vector of k entries, indexed by $\mathscr{X}^{(\ell)}$, in the bottom right block of X and let $Y^{(\ell)} = (Y_{(i_1^{(\ell)}, j_1^{(\ell)})}, \ldots, Y_{(i_k^{(\ell)}, j_k^{(\ell)})})$. By the Azuma-Hoeffding inequality, we have

$$\mathbb{P}\left(\left|\frac{1}{[n^{2\delta}/k]}\sum_{\ell=1}^{[n^{2\delta}/k]}\mathbb{1}_{\{X^{(\ell)}\in\mathcal{A}\}} - \mathbb{P}\left(Y^{(1)}\in\mathcal{A}\right)\right| > \frac{1}{2}\varepsilon\right) \le c' \cdot \frac{\exp(n\log n)}{\exp\left(-\frac{\varepsilon^2}{8}\cdot[n^{2\delta}/k]\right)}$$

for sufficiently large n. Hence, when $k = o\left(\frac{n^{2\delta-1}}{\log n}\right)$, we have

$$\mathbb{P}\left(\left|\frac{1}{[n^{2\delta}/k]}\sum_{\ell=1}^{[n^{2\delta}/k]}\mathbb{1}_{\{X^{(\ell)}\in\mathcal{A}\}} - \mathbb{P}\left(Y^{(1)}\in\mathcal{A}\right)\right| > \frac{1}{2}\varepsilon\right) = o(1).$$

Since all $X^{(\ell)}$'s have the same distribution for all $1 \le \ell \le [n^{2\delta}/k]$, we have

$$\left|\mathbb{P}(X^{(1)}\in\mathcal{A}) - \mathbb{P}(Y^{(1)}\in\mathcal{A})\right| \le \frac{1}{2}\varepsilon + o(1).$$

Again, by (3.2), we have

$$\begin{aligned} d_{TV}\left(Y^{(1)}, V_k\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) &\leq k \cdot d_{TV}\left(\mathrm{Ber}(z_{11}), \mathrm{Ber}\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) \\ &= 2k \cdot \left| z_{11} - \frac{B^2(1-C)}{B^2-2B+1/C} \right| \\ &= k \cdot O(n^{\delta-1}). \end{aligned}$$

Hence, when $k = o\left(n^{1-\delta}\right)$, we have $d_{TV}\left(Y^{(1)}, V_k\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) = o(1)$. Notice that

$$\begin{cases} 1 \ll \frac{n^{2\delta-1}}{\log n} \ll n^{1-\delta} & \text{if } \frac{1}{2} < \delta \leq \frac{2}{3}, \\ 1 \ll n^{1-\delta} \ll \frac{n^{2\delta-1}}{\log n} & \text{if } \frac{2}{3} < \delta < 1. \end{cases}$$

Hence, by triangle inequality, when

$$k = k(n) = \begin{cases} o\left(\frac{n^{2\delta-1}}{\log n}\right) & \text{if } \frac{1}{2} < \delta \leq \frac{2}{3}, \\ o(n^{1-\delta}) & \text{if } \frac{2}{3} < \delta < 1, \end{cases}$$

we have

$$d_{TV}\left((X_{11}, \ldots, X_{1k}), V_k\left(\frac{B^2(1-C)}{B^2-2B+1/C}\right)\right) \to 0.$$

This finishes the proof of (2).

Finally, we prove (3). Likewise, let

$$\mathscr{X}^{(\ell)} = \{(i_1^{(\ell)}, j_1^{(\ell)}), \ldots, (i_k^{(\ell)}, j_k^{(\ell)}) : 1 \leq i_r^{(\ell)}, j_r^{(\ell)} \leq [n^{1+\delta}]\}$$

be a k-subset of indices of *bottom left* block with $\mathscr{X}^{(\ell)} \cap \mathscr{X}^{(\ell')} = \emptyset$ if $\ell \neq \ell'$ and $1 \leq \ell \leq [n^{1+\delta}/k]$ and let $X^{(\ell)} = (X_{(i_1^{(\ell)}, j_1^{(\ell)})}, \ldots, X_{(i_k^{(\ell)}, j_k^{(\ell)})})$ be the random vector of k entries, indexed by $\mathscr{X}^{(\ell)}$, within the *bottom left* block of X.

Let $Y^{(\ell)} = (Y_{(i_1^{(\ell)}, j_1^{(\ell)})}, \ldots, Y_{(i_k^{(\ell)}, j_k^{(\ell)})})$, where $Y = (Y_{ij})$ is the matrix of independent Bernoulli random variables with $Y_{ij} \sim \mathrm{Ber}(z_{ij})$. By the Azuma-Hoeffding inequality,

$$\mathbb{P}\left(\left|\frac{1}{[n^{1+\delta}/k]} \sum_{\ell=1}^{[n^{1+\delta}/k]} \mathbb{1}_{\{X^{(\ell)} \in \mathcal{A}\}} - \mathbb{P}\left(Y^{(1)} \in \mathcal{A}\right)\right| > \frac{1}{2}\varepsilon\right) \leq c' \frac{\exp(n \log n)}{\exp\left(-\frac{\varepsilon^2}{8} \cdot [n^{1+\delta}/k]\right)}$$

for sufficiently large n. Hence, when $k = o\left(\frac{n^{\delta}}{\log n}\right)$,

$$\mathbb{P}\left(\left|\frac{1}{[n^{1+\delta}/k]} \sum_{\ell=1}^{[n^{1+\delta}/k]} \mathbb{1}_{\{X^{(\ell)} \in \mathcal{A}\}} - \mathbb{P}\left(Y^{(1)} \in \mathcal{A}\right)\right| > \frac{1}{2}\varepsilon\right) = o(1).$$

Since all $X^{(\ell)}$'s have the same distribution for $1 \leq \ell \leq [n^{1+\delta}/k]$,

$$\left|\mathbb{P}(X^{(1)} \in \mathcal{A}) - \mathbb{P}(Y^{(1)} \in \mathcal{A})\right| \leq \frac{1}{2}\varepsilon + o(1).$$

By (3.2),

$$\begin{aligned} d_{TV}\left(Y^{(1)}, V_k(BC)\right) &\leq k \cdot d_{TV}\left(\mathrm{Ber}(z_{1,n+1}), \mathrm{Ber}(BC)\right) \\ &= 2k \cdot |z_{1,n+1} - BC| \\ &= k \cdot O(n^{\delta-1}). \end{aligned}$$

Hence, when $k = o\left(n^{1-\delta}\right)$, we have $d_{TV}\left(Y^{(1)}, V_k\left(BC\right)\right) = o(1)$. Furthermore, notice that

$$\begin{cases} 1 \ll \frac{n^\delta}{\log n} \ll n^{1-\delta} & \text{if } 0 < \delta \leq \frac{1}{2}, \\ 1 \ll n^{1-\delta} \ll \frac{n^\delta}{\log n} & \text{if } \frac{1}{2} < \delta < 1. \end{cases}$$

Hence, by triangle inequality, when

$$k = k(n) = \begin{cases} o\left(\frac{n^\delta}{\log n}\right) & \text{if } 0 < \delta \leq \frac{1}{2}, \\ o(n^{1-\delta}) & \text{if } \frac{1}{2} < \delta < 1, \end{cases}$$

we have

$$d_{TV}\left((X_{1,[n^\delta]+1}, \ldots, X_{1,[n^\delta]+k}), V_k(BC)\right) \to 0.$$

This finishes the proof of (3). $\square$

A direct consequence of Theorem 1.4 is the below Corollary 3.4. To state the results, let $J = J(n, \delta, B, C) = (J_{ij})$ be the matrix of independent Bernoulli random variables such that

$$J_{ij} \sim \begin{cases} \mathrm{Ber}(C) & \text{if } 1 + [n^\delta] \leq i, j \leq n + [n^\delta], \\ \mathrm{Ber}(BC) & \text{if } 1 \leq i \leq [n^\delta], [n^\delta] + 1 \leq j \leq [n^\delta] + n, \\ \mathrm{Ber}(BC) & \text{if } 1 \leq j \leq [n^\delta], [n^\delta] + 1 \leq i \leq [n^\delta] + n, \\ \mathrm{Ber}\left(\frac{B^2(1-C)}{B^2 - 2B + 1/C}\right) & \text{if } 1 \leq i, j \leq [n^\delta]. \end{cases}$$

Corollary 3.4. *Let $(i_1, j_1), \ldots, (i_L, j_L)$ be a fixed sequence of pairs of positive integers and let $\alpha_1, \ldots, \alpha_L$ be a fixed sequence of positive integers. Under the exact same conditions as Theorem 1.4, we have*

$$\mathbb{E}\left[\prod_{k=1}^{L} X_{i_k,j_k}^{\alpha_k}\right] \to \mathbb{E}\left[\prod_{k=1}^{L} J_{i_k,j_k}^{\alpha_k}\right] \tag{3.4}$$

if $1 \leq i_k, j_k \leq [n^\delta]$ or $[n^\delta] + 1 \leq i, j \leq [n^\delta] + n$ or $1 \leq i \leq [n^\delta], [n^\delta] + 1 \leq j \leq [n^\delta] + n$ or $1 \leq j \leq [n^\delta], [n^\delta] + 1 \leq i \leq [n^\delta] + n$. In other words, the $(i_k, j_k)'s$ are in the same block for all $1 \leq k \leq L$.

4. Proof of Theorem 1.5 and 1.6

In this section, we prove the Theorem 1.5 and 1.6. Notice that similar results have been obtained in the non-negative integer case Dittmer et al. (2020). We first obtain the explicit convergence rate for (3.4). Notice that

$$\begin{aligned}
&\left|\mathbb{E}\left[\prod_{k=1}^{L} X_{i_k,j_k}^{\alpha_k}\right] - \mathbb{E}\left[\prod_{k=1}^{L} J_{i_k,j_k}^{\alpha_k}\right]\right| \\
&= \left|\mathbb{P}\left(\prod_{k=1}^{L} X_{i_k,j_k}^{\alpha_k} = 1\right) - \mathbb{P}\left(\prod_{k=1}^{L} J_{i_k,j_k}^{\alpha_k} = 1\right)\right| \\
&= \left|\mathbb{P}\left(\prod_{k=1}^{L} X_{i_k,j_k} = 1\right) - \mathbb{P}\left(\prod_{k=1}^{L} J_{i_k,j_k} = 1\right)\right| \\
&\leq d_{TV}\left(\prod_{k=1}^{L} X_{i_k,j_k}, \prod_{k=1}^{L} J_{i_k,j_k}\right) \\
&\leq d_{TV}\left(\prod_{k=1}^{L} X_{i_k,j_k}, \prod_{k=1}^{L} Y_{i_k,j_k}\right) + d_{TV}\left(\prod_{k=1}^{L} Y_{i_k,j_k}, \prod_{k=1}^{L} J_{i_k,j_k}\right).
\end{aligned}$$

Recall that $Y = (Y_{ij})$ is the matrix of independent Bernoulli random variables with each Y_{ij} has mean z_{ij}. The matrix $Z = (z_{ij})$ is the typical table. By symmetry and (3.2), we have

$$d_{TV}\left(\prod_{k=1}^{L} Y_{i_k,j_k}, \prod_{k=1}^{L} J_{i_k,j_k}\right) \leq L \cdot d_{TV}\left(Y_{i_1,j_1}, J_{i_1,j_1}\right) = O(n^{\delta-1}).$$

For the $d_{TV}\left(\prod_{k=1}^{L} X_{i_k,j_k}, \prod_{k=1}^{L} Y_{i_k,j_k}\right)$, it can be shown that

$$d_{TV}\left(\prod_{k=1}^{L} X_{i_k,j_k}, \prod_{k=1}^{L} Y_{i_k,j_k}\right) = O\left(n^{-\eta(\delta)+\varepsilon}\right), \tag{4.1}$$

where

$$\eta(\delta) = \begin{cases} \frac{1}{2} & \text{if all the } (i_k, j_k)'s \text{ are in the bottom right block,} \\ \delta - \frac{1}{2} & \text{if all the } (i_k, j_k)'s \text{ are in the top left block,} \\ \frac{\delta}{2} & \text{otherwise.} \end{cases} \tag{4.2}$$

The proof of (4.1) is similar to that of Lemma 3.3 and Dittmer et al. (2020, Theorem 6.1) so the details are omitted. Hence, we have

$$\left|\mathbb{E}\left[\prod_{k=1}^{L} X_{i_k,j_k}^{\alpha_k}\right] - \mathbb{E}\left[\prod_{k=1}^{L} J_{i_k,j_k}^{\alpha_k}\right]\right| = O\left(n^{\delta-1} + n^{-\eta(\delta)+\varepsilon}\right), \tag{4.3}$$

where $\eta(\delta)$ is defined in (4.2).

Next, we prove Theorem 1.5. Note that the proof of Theorem 1.6 follows the exact same reasoning so we will not provide any details here.

Proof of Theorem 1.5: Let $\overline{X}_{1,[n^\delta]+k} = X_{1,[n^\delta]+k} - BC$ for all $1 \leq k \leq n$ and let

$$\begin{aligned} \overline{S}_{n,\delta}(B,C) &:= X_{1,[n^\delta]+1} + \ldots + X_{1,[n^\delta]+n} - BCn \\ &= \overline{X}_{1,[n^\delta]+1} + \ldots + \overline{X}_{1,[n^\delta]+n}. \end{aligned}$$

By Markov's inequality, we have

$$\begin{aligned} \mathbb{P}\left(\overline{S}_{n,\delta}(B,C) > t\right) &\leq \frac{1}{t^2}\mathbb{E}\left[\left(\sum_{k=1}^{n} \overline{X}_{1,k+[n^\delta]}\right)^2\right] \\ &= \frac{1}{t^2}\mathbb{E}\left[\sum_{k=1}^{n} \overline{X}_{1,k+[n^\delta]}^2 + 2\sum_{[n^\delta]+1 \leq k_1 \neq k_2 \leq [n^\delta]+n} \overline{X}_{1k_1}\overline{X}_{1k_2}\right] \\ &= \frac{1}{t^2}\left\{n\mathbb{E}\left[\overline{X}_{1,n+1}^2\right] + n(n-1)\mathbb{E}\left[\overline{X}_{1,n+1}\overline{X}_{1,n+2}\right]\right\}. \end{aligned}$$

By (4.3), we have

$$\mathbb{E}\left[\overline{X}_{1,n+1}^2\right] = BC - B^2C^2 + O\left(n^{\delta-1} + n^{-\frac{\delta}{2}+\varepsilon}\right)$$

and

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

Therefore, when $0 < \delta \leq \frac{1}{2}$, we have

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

Hence, for all $\xi, \varepsilon > 0$, there exists some constant $c' > 0$ such that

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

for sufficiently large n. If we choose $0 < \xi < \frac{\delta}{4}$, then for some constants $c'', \xi' > 0$,

$$\mathbb{P}\left(\frac{\overline{S}_{n,\delta}(B,C)}{n} > n^{-\xi}\right) \le c'' n^{-\xi'}$$

for sufficiently large n. For any sequence $(n_k)_{k\ge 1}$ with $n_k \to \infty$ as $k \to \infty$, there exists a subsequence $(n_{k_r})_{r\ge 1}$ with $n_{k_r} \to \infty$ as $r \to \infty$ such that

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

By the Borel-Cantelli Lemma, we have

$$\frac{\overline{S}_{n_{k_r},\delta}(B,C)}{n_{k_r}} \to 0 \qquad \text{as } r \to \infty$$

almost surely. Consequently,

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

almost surely. This completes the proof. □

Acknowledgements

The author expresses gratitude to an anonymous referee for his valuable feedback, which contributed to the correction of errors and the overall improvement of the manuscript. The author would also like to thank Robin Pemantle for many helpful discussions.

References

Artstein, S., Ball, K. M., Barthe, F., and Naor, A. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.*, **17** (4), 975–982 (2004). MR2083473.

Barvinok, A. On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries. *Adv. Math.*, **224** (1), 316–339 (2010a). MR2600999.

Barvinok, A. What does a random contingency table look like? *Combin. Probab. Comput.*, **19** (4), 517–539 (2010b). MR2647491.

Barvinok, A. Matrices with prescribed row and column sums. *Linear Algebra Appl.*, **436** (4), 820–844 (2012). MR2890890.

Barvinok, A. and Hartigan, J. A. Maximum entropy Gaussian approximations for the number of integer points and volumes of polytopes. *Adv. in Appl. Math.*, **45** (2), 252–289 (2010). MR2646125.

Barvinok, A. and Hartigan, J. A. An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. *Trans. Amer. Math. Soc.*, **364** (8), 4323–4368 (2012). MR2912457.

Borel, E. *Introduction géométrique à quelques théories physiques*. Gauthier-Villars (1914).

Canfield, E. R. and McKay, B. D. Asymptotic enumeration of integer matrices with large equal row and column sums. *Combinatorica*, **30** (6), 655–680 (2010). MR2789732.

Chatterjee, S., Diaconis, P., and Sly, A. Properties of uniform doubly stochastic matrices. *ArXiv Mathematics e-prints* (2010). arXiv: 1010.6136.

Diaconis, P. and Gangolli, A. Rectangular arrays with fixed margins. In *Discrete probability and algorithms (Minneapolis, MN, 1993)*, volume 72 of *IMA Vol. Math. Appl.*, pp. 15–41. Springer, New York (1995). MR1380519.

Diaconis, P. and Shahshahani, M. On the eigenvalues of random matrices. *J. Appl. Probab.*, **31A**, 49–62 (1994). Studies in applied probability. MR1274717.

Dittmer, S., Lyu, H., and Pak, I. Phase transition in random contingency tables with non-uniform margins. *Trans. Amer. Math. Soc.*, **373** (12), 8313–8338 (2020). MR4177260.

Good, I. J. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.*, **34**, 911–934 (1963). MR150880.

Good, I. J. On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Ann. Statist.*, **4** (6), 1159–1189 (1976). MR428568.

Good, I. J. and Crook, J. F. The enumeration of arrays and a generalization related to contingency tables. *Discrete Math.*, **19** (1), 23–45 (1977). MR541011.

Lang, S. *A First Course in Calculus*. Undergraduate Texts in Mathematics. Springer New York, NY, 5 edition (1998). ISBN 978-0-387-96201-6. DOI: 10.1007/978-1-4419-8532-3.

Lyu, H. and Pak, I. On the number of contingency tables and the independence heuristic. *Bull. Lond. Math. Soc.*, **54** (1), 242–255 (2022). MR4403278.

Macdonald, I. G. *Symmetric functions and orthogonal polynomials*, volume 12 of *University Lecture Series*. American Mathematical Society, Providence, RI (1998). ISBN 0-8218-0770-6. MR1488699.

Nguyen, H. H. Random doubly stochastic matrices: the circular law. *Ann. Probab.*, **42** (3), 1161–1196 (2014). MR3189068.

Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.*, **27**, 379–423, 623–656 (1948). MR26286.

Stanley, R. P. *Enumerative combinatorics. Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition (2012). ISBN 978-1-107-60262-5. MR2868112.

van Lint, J. H. and Wilson, R. M. *A course in combinatorics*. Cambridge University Press, Cambridge, second edition (2001). ISBN 0-521-00601-5. MR1871828.

Wu, D. Asymptotic enumeration of binary contingency tables and comparison with independence heuristic. *Contrib. Math.*, **8**, 16–23 (2023a). DOI: 10.47443/cm.2023.037.

Wu, D. Asymptotic Properties of Random Contingency Tables with Uniform Margin. *J. Theoret. Probab.*, **36** (4), 2066–2092 (2023b). MR4665839.