

Research Article

Asymptotic enumeration of binary contingency tables and comparison with independence heuristic

Da Wu*

Department of Mathematics, David Rittenhouse Laboratory, University of Pennsylvania, Philadelphia, Pennsylvania, USA

(Received: 27 June 2023. Received in revised form: 30 July 2023. Accepted: 1 August 2023. Published online: 2 August 2023.)

© 2023 the author. This is an open access article under the CC BY (International 4.0) license (www.creativecommons.org/licenses/by/4.0/).

Abstract

For parameters n, δ, B, C , a sharp asymptotic formula for the number of $(n + \lfloor n^\delta \rfloor)^2$ -dimensional binary contingency tables with non-uniform margins taking values of $\lfloor BCn \rfloor$ and $\lfloor Cn \rfloor$ is obtained. The obtained sharp asymptotics is compared with the classical *independence heuristic estimate* and it is proved that the independence heuristic overestimates by a factor of $e^{\Theta(n^{2\delta})}$. The comparison is based on the analysis of the *correlation ratio*. An explicit bound for the constant in Θ is also obtained.

Keywords: binary contingency tables; maximum entropy principle; independence heuristic.

2020 Mathematics Subject Classification: 05A16.

1. Introduction

Overview

This paper studies the asymptotic enumeration of *binary contingency tables* and its connection with the classical *independence heuristic* introduced by I. J. Good 70 years ago [4]. A binary contingency table is a set of 0-1 matrices with fixed row and column sums. Let $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{c} = (c_1, \dots, c_n)$ be two positive integer vectors with the same total sum of entries, i.e., $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j = N$. Let

$$\mathcal{M}(\mathbf{r}, \mathbf{c}) = \left\{ X = (X_{ij}) : \sum_{k=1}^n X_{ik} = r_i, \sum_{k=1}^m X_{kj} = c_j \text{ for all } 1 \leq i \leq m, 1 \leq j \leq n \right\}$$

be the set of binary contingency tables with row margin \mathbf{r} and column margin \mathbf{c} . Since $X_{ij} \in \{0, 1\}$, it is easy to see that $r_i \leq n, c_j \leq m$ for all i, j . A binary contingency table has close connections with bipartite graphs with fixed degree sequence; e.g., see [8] for a historical review. It also arises as the structural constants in the ring of symmetric functions and representation theory of general linear groups (see [6]).

Estimating the cardinality of $\mathcal{M}(\mathbf{r}, \mathbf{c})$ is a fundamental problem in analytic combinatorics; for instance, see [1, 4]. At the very beginning, we have the following effortless estimate based on the so-called *independence heuristic*. Precisely speaking, fix $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{c} = (c_1, \dots, c_n)$ with $N = r_1 + \dots + r_m = c_1 + \dots + c_n$, and let \mathcal{M}_N be the set of $m \times n$ 0-1 matrices with total sum of entries N . Let X be the uniform sample from \mathcal{M}_N and consider the following two events:

$$\mathcal{R}_{\mathbf{r}} = \{X \text{ has row sum } \mathbf{r}\} \quad \text{and} \quad \mathcal{R}_{\mathbf{c}} = \{X \text{ has column sum } \mathbf{c}\}.$$

It follows from the definition that

$$\mathbb{P}(\mathcal{R}_{\mathbf{r}}) = \frac{|\mathcal{R}_{\mathbf{r}}|}{|\mathcal{M}_N|} \quad \text{and} \quad \mathbb{P}(\mathcal{R}_{\mathbf{c}}) = \frac{|\mathcal{R}_{\mathbf{c}}|}{|\mathcal{M}_N|}.$$

Assume that $\mathcal{R}_{\mathbf{r}}$ and $\mathcal{R}_{\mathbf{c}}$ are independent, then

$$\frac{|\mathcal{R}_{\mathbf{r}}|}{|\mathcal{M}_N|} \cdot \frac{|\mathcal{R}_{\mathbf{c}}|}{|\mathcal{M}_N|} = \mathbb{P}(\mathcal{R}_{\mathbf{r}}) \mathbb{P}(\mathcal{R}_{\mathbf{c}}) = \mathbb{P}(\mathcal{R}_{\mathbf{r}} \cap \mathcal{R}_{\mathbf{c}}) = \frac{|\mathcal{M}(\mathbf{r}, \mathbf{c})|}{|\mathcal{M}_N|}.$$

Therefore,

$$|\mathcal{M}(\mathbf{r}, \mathbf{c})| = \frac{|\mathcal{R}_{\mathbf{r}}| \cdot |\mathcal{R}_{\mathbf{c}}|}{|\mathcal{M}_N|} = \binom{mn}{N}^{-1} \prod_{i=1}^m \binom{n}{r_i} \prod_{j=1}^n \binom{m}{c_j}. \quad (1)$$

*E-mail address: dawu@math.upenn.edu

We call

$$\mathcal{I}(\mathbf{r}, \mathbf{c}) = \binom{mn}{N}^{-1} \prod_{i=1}^m \binom{n}{r_i} \prod_{j=1}^n \binom{m}{c_j}. \tag{2}$$

the *independence heuristic estimate* corresponding to margins \mathbf{r} and \mathbf{c} .

In 1969, O’Neil [7] solved the problem for the case of large sparse binary matrices with $r_i, c_j \leq (\log n)^{\frac{1}{4}-\varepsilon}$ for all $1 \leq i, j \leq n$. In 2010, Barvinok [1] used *permanents* and the *van der Waerden bound* for doubly-stochastic matrix to obtain an estimate for generic \mathbf{r} and \mathbf{c} . This estimate has recently been improved in [3] using the techniques of Lorentzian polynomials. Later on, Barvinok and Hartigan [2] used the *Maximal Entropy Principle* and local central limit theorem to obtain a more precise asymptotic formula under certain regularity conditions.

Our main focus is to first derive the asymptotics of $|\mathcal{M}(\mathbf{r}, \mathbf{c})|$ when both \mathbf{r} and \mathbf{c} take two different linear values of n ; see the precise definition in the next subsection. After that, we compared the results with the independence heuristic estimate (2) and showed that the independence heuristic estimate leads to a large overestimate by a factor of $e^{\Theta(n^{2\delta})}$. Our derivation of asymptotics follows closely the spirit of [5] and is mainly based on Barvinok’s asymptotic formula in [1] and the author’s recent work [9] on the limiting distribution of *Random Binary Contingency Tables*. Similar asymptotics for the uniform margin case can also be obtained with the same techniques and the limiting distribution derived in [10].

Setup and statements of main results

For $0 < \delta < 1, 0 < B \leq \frac{1}{C}$ and $0 < C < 1$, let

$$\tilde{\mathbf{r}} = \tilde{\mathbf{c}} := (\underbrace{\lfloor BCn \rfloor, \dots, \lfloor BCn \rfloor}_{\lfloor n^\delta \rfloor \text{ entries}}, \underbrace{\lfloor Cn \rfloor, \dots, \lfloor Cn \rfloor}_n) \in \mathbb{N}^{\lfloor n^\delta \rfloor + n},$$

and let

$$\mathcal{M}_{n,\delta}(B, C) := \mathcal{M}(\tilde{\mathbf{r}}, \tilde{\mathbf{c}}).$$

Namely, $\mathcal{M}_{n,\delta}(B, C)$ is the set of $(\lfloor n^\delta \rfloor + n)^2$ -dimensional binary matrices whose first $\lfloor n^\delta \rfloor$ rows and columns have sum $\lfloor BCn \rfloor$ and remaining n rows and columns have sum $\lfloor Cn \rfloor$.

The first main result of this paper is the sharp asymptotics of $|\mathcal{M}_{n,\delta}(B, C)|$.

Theorem 1.1. *Fix $0 < \delta < 1, n \geq 1, 0 < C < \frac{3}{4}$ and $0 < B \leq \frac{1}{C}$, and let*

$$f(x) := x \log \frac{1}{x} + (1 - x) \log \frac{1}{1 - x}.$$

Then

$$\begin{aligned} \log |\mathcal{M}_{n,\delta}(B, C)| &= f(C)n^2 + \left[2f(BC) - (BC) \log \left(\frac{1 - C}{C} \right) \right] n^{1+\delta} \\ &\quad + \left[f(z_{11}^*) + z_{11}^* \log \left(\frac{1 - C}{C} \cdot \frac{(BC)^2}{(1 - BC)^2} \right) - \frac{B^2C}{2(1 - C)} \right] n^{2\delta} + O(n^{3\delta-1} + n \log n), \end{aligned}$$

where

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

Theorem 1.1 is proved in Section 2. For now, we remark that the proof is based on the *Maximum Entropy Principle* and the function $f(x)$ is Shannon-Boltzmann entropy of Bernoulli random variable with mean x ; the coefficients in front of $n^2, n^{1+\delta}, n^{2\delta}$ all come from the Bernoulli entropy.

Next, we study the relationship between $|\mathcal{M}_{n,\delta}(B, C)|$ and its corresponding independence heuristic. Recall that for margins $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{c} = (c_1, \dots, c_n)$, the *independence heuristic estimate* is the following quantity,

$$\mathcal{I}(\mathbf{r}, \mathbf{c}) = \binom{mn}{N}^{-1} \prod_{i=1}^m \binom{n}{r_i} \prod_{j=1}^n \binom{m}{c_j},$$

where $N = \sum_{i=1}^m r_i = \sum_{j=1}^n c_j$ is the total sum of entries. We denote

$$\mathcal{I}_{n,\delta}(B, C) := \mathcal{I}(\tilde{\mathbf{r}}, \tilde{\mathbf{c}}).$$

To study the relation between $\mathcal{I}_{n,\delta}(B, C)$ and $|\mathcal{M}_{n,\delta}(B, C)|$, we consider their *Correlation Ratio* $\rho_{n,\delta}(B, C)$. It is defined as

$$\rho_{n,\delta}(B, C) := \frac{|\mathcal{M}_{n,\delta}(B, C)|}{\mathcal{I}_{n,\delta}(B, C)}.$$

Our next main result is on the asymptotic behavior of $\rho_{n,\delta}(B, C)$.

Theorem 1.2. Fix $0 < \delta < 1$, $n \geq 1$, $0 < C < \frac{3}{4}$ and $0 < B \leq \frac{1}{C}$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \rho_{n,\delta}(B, C) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{n^{1+\delta}} \log \rho_{n,\delta}(B, C) = 0.$$

Furthermore, let

$$\Delta_{B,C} := \lim_{n \rightarrow \infty} \frac{1}{n^{2\delta}} \log \rho_{n,\delta}(B, C),$$

and we have the following explicit formula for $\Delta_{B,C}$:

$$\Delta_{B,C} = 1 - \frac{B^2C - 2BC + 1}{1 - C} - \log \left(\frac{1 - C}{B^2C - 2BC + 1} \right).$$

Moreover,

$$0 \geq \Delta_{B,C} > \begin{cases} -\frac{1}{C} + \log \frac{1}{C} + 1 & \text{when } 0 < C \leq \frac{1}{2}, \\ -\frac{1}{1-C} + \log \frac{1}{1-C} + 1 & \text{when } \frac{1}{2} < C < \frac{3}{4}, \end{cases}$$

with $\Delta_{B,C} = 0$ if and only if $B = 1$.

The behavior of $\Delta_{B,C}$ tells us that the independence heuristic overestimates the number of tables in $\mathcal{M}_{n,\delta}(B, C)$. This matches Barvinok’s arguments on *cloned margins*; see [1] for details. Probabilistically speaking, the events

$$\mathcal{R}_{n,\delta}(B, C) = \{0\text{-1 matrices has row sums } \tilde{\mathbf{r}}\}$$

and

$$\mathcal{C}_{n,\delta}(B, C) = \{0\text{-1 matrices has column sums } \tilde{\mathbf{c}}\}$$

are asymptotically *negatively correlated* instead of being asymptotically independent. The $\Delta_{B,C}$ quantifies how far they are away from being independent. In fact, when $B = 1$, i.e., all row sums and columns sums are equal, the independence heuristic provides the best estimate. As B moves away from 1, meaning that the margins become less and less uniform, the independence heuristic overestimates by a factor of $e^{\Theta(n^{2\delta})}$.

It is really interesting for the readers to compare our results with the recent work of [5] on non-negative integer case. When the contingency tables are non-negative integer valued, the independence heuristic leads to a large undercounting, which is opposite to our binary case. The reason behind this phenomenon remains mysterious. It would be nice if we can obtain the intermediate results on the contingency tables whose entries take values from $\{0, 1, \dots, k\}$ for finite k .

2. Proof of Theorem 1.1

Let $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{N}^m$ and $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{N}^n$ be two positive integer vectors with total sum of entries. The *binary transportation polytope* $\mathcal{P}(\mathbf{r}, \mathbf{c})$ is defined as

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

Barvinok introduced the following notion of *Typical Table* in [1].

Definition 2.1 (Typical Table). Let $\mathbf{r} = (r_1, \dots, r_m)$, $\mathbf{c} = (c_1, \dots, c_n)$ be two positive integer vectors with total sum of entries. For each $X = (X_{ij}) \in \mathcal{P}(\mathbf{r}, \mathbf{c})$, let

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

where

$$f(x) = x \log \frac{1}{x} + (1 - x) \log \frac{1}{1 - x}$$

for $x \in (0, 1)$. The *Typical Table* $Z = (z_{ij})$ is defined as the unique maximizer of g on $\mathcal{P}(\mathbf{r}, \mathbf{c})$.

Remark 2.1. The function g is strictly concave so it attains a unique maximum in the interior of the binary transportation polytope. Hence typical table is well-defined.

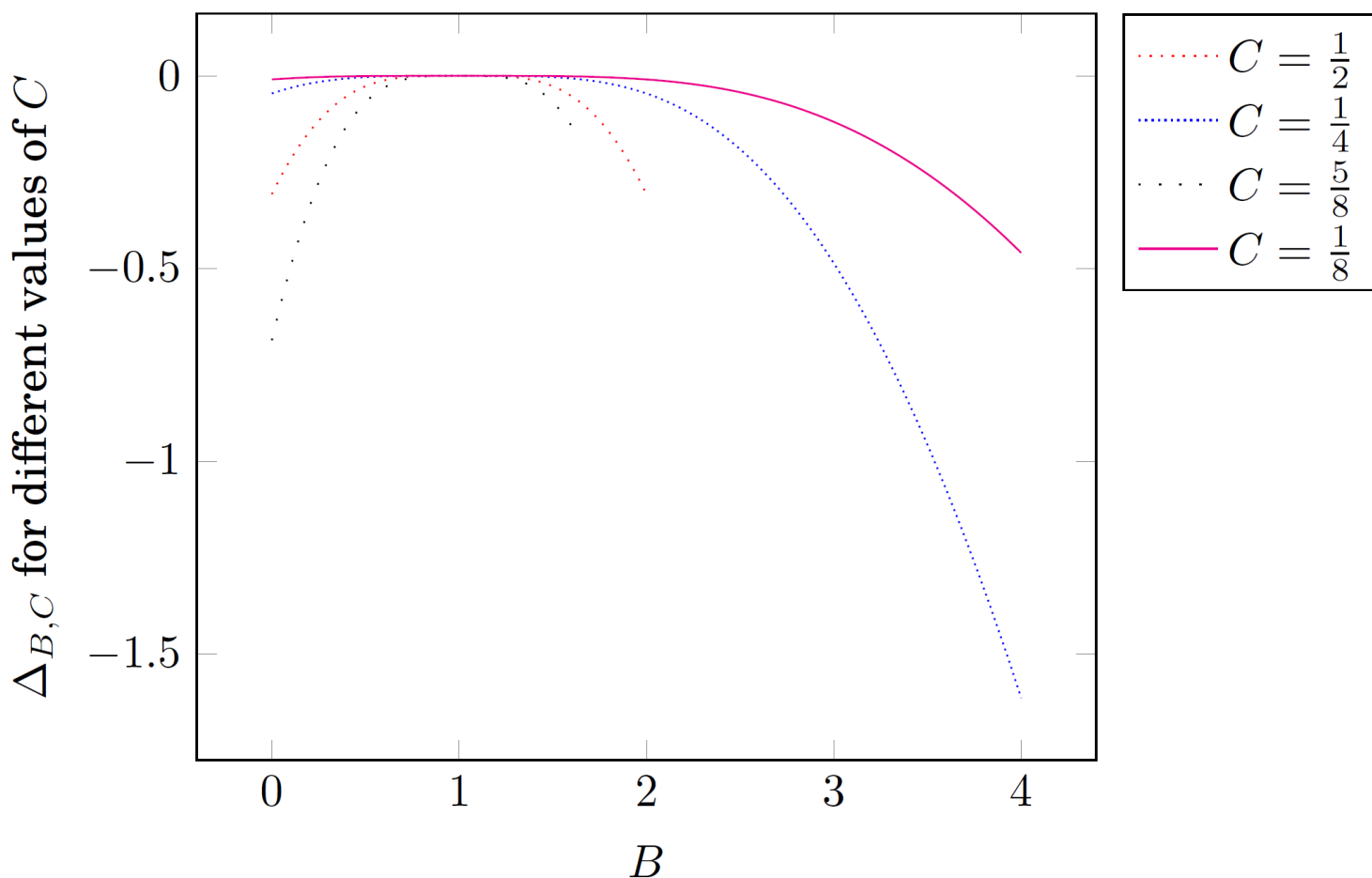


Figure 1: Plot of $\Delta_{B,C} = \lim_{n \rightarrow \infty} \frac{1}{n^{2\delta}} \log \rho_{n,\delta}(B, C)$ as a function of $B \in [0, 4]$ for 4 different fixed values of C . For each $0 < C < \frac{3}{4}$, B is allowed to take value from 0 to $\frac{1}{C}$.

Remark 2.2. The function $f(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}$ is the Shannon-Boltzmann entropy of Bernoulli random variable with mean x . Therefore, $g(X)$ can be viewed as the Bernoulli entropy of X and Z is the maximal entropy matrix on the polytope $\mathcal{P}(\mathbf{r}, \mathbf{c})$.

Remark 2.3. By symmetry, two entries in a typical table are equal if they have the same margin conditions, i.e., $z_{ij} = z_{i'j'}$ if $r_i = r_{i'}, c_j = c_{j'}$.

It turns out that typical table has close connection to the cardinality of $\mathcal{M}(\mathbf{r}, \mathbf{c})$. The following theorem proved by Barvinok in [1] plays a key role in our proof.

Theorem 2.1. Fix row margins $\mathbf{r} = (r_1, \dots, r_m)$ and column margins $\mathbf{c} = (c_1, \dots, c_n)$ and let $Z = (z_{ij})$ be the typical table associated with $\mathcal{P}(\mathbf{r}, \mathbf{c})$. There exists some absolute constant $\gamma > 0$ such that

$$(mn)^{-\gamma(m+n)} e^{g(Z)} \leq |\mathcal{M}(\mathbf{r}, \mathbf{c})| \leq e^{g(Z)}. \tag{3}$$

Next, by [9, (2.3)] and [9, Lemma 2.4], we have the following lemma on the asymptotics of entries of typical table $Z = (z_{ij})$.

Lemma 2.1. Fix $0 \leq \delta < 1$, $0 < C < \frac{3}{4}$ and $0 < B \leq \frac{1}{C}$. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B, C)$. Then there exists constants $\gamma_1(B, C)$ and $\gamma_2(B, C)$ such that the followings hold:

1. $\left| z_{11} - \frac{B^2(1-C)}{B^2-2B+1/C} \right| \leq \gamma_1(B, C)n^{\delta-1}$,
2. $|z_{1,n+1} - BC| \leq \gamma_2(B, C)n^{\delta-1}$,
3. $|z_{n+1,n+1} - C| \leq BCn^{\delta-1}$.

Next, we derive the asymptotics of $g(Z)$, which is the entropy for the typical table. Our proof is almost identical to the [5, Proposition 3.4], except that we plug in different limits of the typical table. Notice that in the binary case, there is no sharp phase transition for the value of B with respect to C ; see [9, Remark 1.4] for more discussions on this aspect.

Proposition 2.1. Fix $0 < \delta < 1$, $0 < C < \frac{3}{4}$ and $0 < B \leq \frac{1}{C}$. Let $Z = (z_{ij})$ be the typical table for $\mathcal{M}_{n,\delta}(B, C)$. Then

$$\begin{aligned} g(Z) &= n^2 f(z_{n+1,n+1}) + 2n[n^\delta]f(z_{1,n+1}) + [n^\delta]^2 f(z_{11}) \\ &= f(C)n^2 + \left[2f(BC) - (BC) \log \left(\frac{1-C}{C} \right) \right] n^{1+\delta} \\ &\quad + \left[f(z_{11}^*) + z_{11}^* \log \left(\frac{1-C}{C} \cdot \frac{(BC)^2}{(1-BC)^2} \right) - \frac{B^2 C}{2(1-C)} \right] n^{2\delta} + O(n^{3\delta-1}) + O(n), \end{aligned}$$

where

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

Proof. By symmetry of typical table and marginal condition,

$$\begin{cases} [n^\delta]z_{11} + nz_{1,n+1} = [BCn] \\ [n^\delta]z_{1,n+1} + nz_{n+1,n+1} = [Cn] \end{cases}.$$

Therefore,

$$\begin{cases} ([n^\delta]/n)z_{11} + z_{1,n+1} = [BCn]/n = BC + O(n^{-1}) \\ ([n^\delta]/n)z_{1,n+1} + z_{n+1,n+1} = [Cn]/n = C + O(n^{-1}) \end{cases}$$

Let $z_{11}^* = \frac{B^2(1-C)}{B^2 - 2B + 1/C}$, then

$$\begin{aligned} C - z_{n+1,n+1} &= z_{1,n+1}([n^\delta]/n) + O(n^{-1}) \\ &= (BC)([n^\delta]/n) - (BC - z_{1,n+1})([n^\delta]/n) + O(n^{-1}) \\ &= (BC)([n^\delta]/n) - z_{11}([n^\delta]/n)^2 + O(n^{-1}) \\ &= (BC)([n^\delta]/n) - z_{11}^*([n^\delta]/n)^2 - (z_{11} - z_{11}^*)([n^\delta]/n)^2 + O(n^{-1}) \\ &= (BC)([n^\delta]/n) - z_{11}^*([n^\delta]/n)^2 + O(n^{3\delta-3} + n^{-1}). \end{aligned}$$

Similarly,

$$\begin{aligned} BC - z_{1,n+1} &= ([n^\delta]/n)z_{11} + O(n^{-1}) \\ &= z_{11}^*([n^\delta]/n) + (z_{11} - z_{11}^*)([n^\delta]/n) + O(n^{-1}) \\ &= z_{11}^*([n^\delta]/n) + O(n^{2\delta-2} + n^{-1}). \end{aligned}$$

Taylor expansion of $f(x)$ around a has the following form,

$$f(x) = f(a) + \log \left(\frac{1-a}{a} \right) (x-a) + \frac{1}{2(a-1)a} (x-a)^2 + O(|x-a|^3).$$

Hence,

$$\begin{aligned} f(z_{n+1,n+1}) &= f(C) - \log \left(\frac{1-C}{C} \right) (BC)([n^\delta]/n) + \log \left(\frac{1-C}{C} \right) z_{11}^*([n^\delta]/n)^2 \\ &\quad + \frac{B^2 C}{2(C-1)}([n^\delta]/n)^2 + O(n^{3\delta-3}) + O(n^{\delta-1}) \\ f(z_{1,n+1}) &= f(BC) - \log \left(\frac{1-BC}{BC} \right) z_{11}^*([n^\delta]/n) + \frac{1}{2(BC-1)BC} (z_{11}^*)^2 ([n^\delta]/n)^2 \\ &\quad + O(n^{2\delta-2}) + O(n^{-1}) \\ f(z_{11}) &= f(z_{11}^*) + O(n^{\delta-1}) + O(n^{-1}). \end{aligned}$$

Therefore, we have that

$$\begin{aligned} g(Z) &= n^2 f(z_{n+1,n+1}) + 2n[n^\delta]f(z_{1,n+1}) + [n^\delta]^2 f(z_{11}) \\ &= f(C)n^2 + \left[2f(BC) - (BC) \log \left(\frac{1-C}{C} \right) \right] n^{1+\delta} \\ &\quad + \left[f(z_{11}^*) + z_{11}^* \log \left(\frac{1-C}{C} \cdot \frac{(BC)^2}{(1-BC)^2} \right) - \frac{B^2C}{2(1-C)} \right] n^{2\delta} + O(n^{3\delta-1}) + O(n). \end{aligned}$$

□

Proof of Theorem 1.1. Notice that for margins $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{c}}$, by Theorem 2.1,

$$g(Z) - \gamma' n \log n \leq \log |\mathcal{M}_{n,\delta}(B, C)| \leq g(Z) \tag{4}$$

Now, Theorem 1.1 follows from Proposition 2.1. □

3. Independence heuristic estimate

Using Stirling formula, we can deduce the asymptotics of *independence heuristic estimation*. Similar asymptotics of independence heuristic for non-negative integer-valued contingency tables can be found in [5, Lemma 4.1].

Lemma 3.1. Fix $0 < \delta < 1$, $0 < C < 1$ and $0 < B \leq \frac{1}{C}$,

$$\begin{aligned} \log \mathcal{I}_{n,\delta}(B, C) &= f(C)n^2 + \left[2f(BC) - BC \log \left(\frac{1-C}{C} \right) \right] n^{1+\delta} \\ &\quad + \left[\frac{B^2C - 4BC + 2C}{2(1-C)} + \log(1-C) - 2 \log(1-BC) \right] n^{2\delta} + O(n^{3\delta-1} + n \log n). \end{aligned}$$

Proof. By Stirling formula, we have that

$$\log \binom{a+b}{a} = (a+b) \log(a+b) - a \log a - b \log b + O(\log(a+b)).$$

Recall that for general margins \mathbf{r} and \mathbf{c} , the independence heuristic estimate takes the form

$$\mathcal{I}(\mathbf{r}, \mathbf{c}) = \binom{mn}{N}^{-1} \prod_{i=1}^m \binom{n}{r_i} \prod_{j=1}^n \binom{m}{c_j}.$$

Therefore,

$$\begin{aligned} \log \mathcal{I}(\mathbf{r}, \mathbf{c}) &= N \log N + (mn - N) \log(mn - N) - \sum_{i=1}^m r_i \log r_i - \sum_{i=1}^m (n - r_i) \log(n - r_i) \\ &\quad - \sum_{j=1}^n c_j \log c_j - \sum_{j=1}^n (m - c_j) \log(m - c_j) + O((m+n) \log(mn)). \end{aligned}$$

In our setup, when $\mathbf{r} = \tilde{\mathbf{r}}$ and $\mathbf{c} = \tilde{\mathbf{c}}$, the dimension is $n + n^\delta$ and the total sum of entries $N = BCn^{1+\delta} + Cn^2$. By Taylor expansion

$$\log(x + y) = \log y + \frac{x}{y} - \frac{x^2}{2y^2} + O\left(\frac{x^3}{y^3}\right),$$

we have that

$$\begin{aligned} \log(N) &= \log(Cn^2 + BCn^{1+\delta}) = \log(Cn^2) + Bn^{\delta-1} - \frac{B^2}{2} n^{2\delta-2} + O(n^{3\delta-3}), \\ \log(mn - N) &= \log((1-C)n^2 + (2-BC)n^{1+\delta} + n^{2\delta}) = \log[(1-C)n^2] \\ &\quad + \left(\frac{2-BC}{1-C}\right) n^{\delta-1} + \frac{-2-2C+4BC-B^2C^2}{2(1-C)^2} n^{2\delta-2} + O(n^{3\delta-3}), \\ \log[(1-C)n + n^\delta] &= \log[(1-C)n] + \frac{n^{\delta-1}}{1-C} - \frac{n^{2\delta-2}}{2(1-C)^2} + O(n^{3\delta-3}), \end{aligned}$$

$$\log [(1 - BC)n + n^\delta] = \log [(1 - BC)n] + \frac{n^{\delta-1}}{1 - BC} - \frac{n^{2\delta-2}}{2(1 - BC)^2} + O(n^{3\delta-3}).$$

Therefore, $\log \mathcal{I}_{n,\delta}(B, C)$ has the following expansion,

$$\begin{aligned} \log \mathcal{I}_{n,\delta}(B, C) &= (Cn^2 + BCn^{1+\delta}) \left[\log(Cn^2) + Bn^{\delta-1} - \frac{B^2}{2}n^{2\delta-2} \right] \\ &+ [(1 - C)n^2 + (2 - BC)n^{1+\delta} + n^{2\delta}] \times \left[\log((1 - C)n^2) + \left(\frac{2 - BC}{1 - C} \right) n^{\delta-1} + \frac{-2 - 2C + 4BC - B^2C^2}{2(1 - C)^2} n^{2\delta-2} \right] \\ &- 2BCn^{1+\delta} \log(BCn) - 2n^\delta [(1 - BC)n + n^\delta] \left[\log((1 - BC)n) + \frac{n^{\delta-1}}{1 - BC} - \frac{n^{2\delta-2}}{2(1 - BC)^2} \right] \\ &- 2Cn^2 \log(Cn) - 2n [(1 - C)n + n^\delta] \cdot \left[\log((1 - C)n) + \frac{n^{\delta-1}}{1 - C} - \frac{n^{2\delta-2}}{2(1 - C)^2} \right] + O(n^{3\delta-1} + n \log n). \end{aligned}$$

After reorganizing terms, we have that

$$\begin{aligned} \log \mathcal{I}_{n,\delta}(B, C) &= f(C)n^2 + \left[2f(BC) - BC \log \left(\frac{1 - C}{C} \right) \right] n^{1+\delta} \\ &+ \left[\frac{B^2C - 4BC + 2C}{2(1 - C)} + \log(1 - C) - 2 \log(1 - BC) \right] n^{2\delta} + O(n^{3\delta-1} + n \log n). \end{aligned}$$

□

Proof of Theorem 1.2. By Theorem 1.1,

$$\begin{aligned} \log |\mathcal{M}_{n,\delta}(B, C)| &= f(C)n^2 + \left[2f(BC) - (BC) \log \left(\frac{1 - C}{C} \right) \right] n^{1+\delta} \\ &+ \left[f(z_{11}^*) + z_{11}^* \log \left(\frac{1 - C}{C} \cdot \frac{(BC)^2}{(1 - BC)^2} \right) - \frac{B^2C}{2(1 - C)} \right] n^{2\delta} + O(n^{3\delta-1} + n \log n). \end{aligned}$$

On the other hand, by Lemma 3.1,

$$\begin{aligned} \log \mathcal{I}_{n,\delta}(B, C) &= f(C)n^2 + \left[2f(BC) - BC \log \left(\frac{1 - C}{C} \right) \right] n^{1+\delta} \\ &+ \left[\frac{B^2C - 4BC + 2C}{2(1 - C)} + \log(1 - C) - 2 \log(1 - BC) \right] n^{2\delta} + O(n^{3\delta-1} + n \log n). \end{aligned}$$

From this, we can easily deduce that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \frac{|\mathcal{M}_{n,\delta}(B, C)|}{\mathcal{I}_{n,\delta}(B, C)} = 0, \quad \lim_{n \rightarrow \infty} \frac{1}{n^{1+\delta}} \log \frac{|\mathcal{M}_{n,\delta}(B, C)|}{\mathcal{I}_{n,\delta}(B, C)} = 0$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\delta}} \log \frac{|\mathcal{M}_{n,\delta}(B, C)|}{\mathcal{I}_{n,\delta}(B, C)} = \Delta_{B,C},$$

where

$$-\Delta_{B,C} = \frac{B^2C - 4BC + 2C}{2(1 - C)} + \log \left(\frac{1 - C}{(1 - BC)^2} \right) - f(z_{11}^*) - z_{11}^* \log \left(\frac{1 - C}{(1 - BC)^2} \right) - z_{11}^* \log(B^2C) + \frac{B^2C}{2(1 - C)}.$$

Notice that

$$1 - z_{11}^* = \frac{(BC - 1)^2}{B^2C - 2BC + 1},$$

By direct computation,

$$\begin{aligned} -\Delta_{B,C} &= \frac{2B^2C - 4BC + 2C}{2(1 - C)} + \frac{(BC - 1)^2}{B^2C - 2BC + 1} \log \left[\frac{1 - C}{B^2C - 2BC + 1} \right] + \frac{B^2(1 - C)}{B^2 - 2B + 1/C} \log \left[\frac{1 - C}{B^2C - 2BC + 1} \right] \\ &= \frac{B^2C - 2BC + 1}{1 - C} + \log \left[\frac{1 - C}{B^2C - 2BC + 1} \right] - 1 \\ &\geq 0, \end{aligned}$$

where the last step is based on the simple fact that $x - \log x \geq 1$ for all $x > 0$. For fixed C , the explicit bound of $\Delta_{B,C}$ follows from the elementary analysis. □

Acknowledgment

The author would like to thank Robin Pemantle for many helpful discussions.

References

- [1] 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.
- [2] A. Barvinok, J. Hartigan, The number of graphs and a random graph with a given degree sequence, *Random Structures Algorithms* **42** (2013) 301–348.
- [3] P. Brändén, J. Leake, I. Pak, Lower bounds for contingency tables via Lorentzian polynomials, *Israel J. Math.* **253** (2023) 43–90.
- [4] I. J. Good, J. F. Crook, The enumeration of arrays and a generalization related to contingency tables, *Discrete Math.* **19** (1977) 23–45.
- [5] H. Lyu, I. Pak, On the number of contingency tables and the independence heuristic, *Bull. Lond. Math. Soc.* **54** (2022) 242–255.
- [6] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Clarendon Press, Oxford, 1998.
- [7] P. E. O’Neil, Asymptotics and random matrices with row-sum and column sum-restrictions, *Bull. Amer. Math. Soc.* **75** (1969) 1276–1282.
- [8] N. Wormald, Asymptotic enumeration of graphs with given degree sequence, *Proceedings of the International Congress of Mathematicians (ICM 2018)*, 2019, 3245–3264.
- [9] D. Wu, On properties of random binary contingency tables with non-uniform margin, *arXiv:2002.12559v4 [math.PR]*, (2022).
- [10] D. Wu, Asymptotic properties of random contingency tables with uniform margin, *J. Theoret. Prob.*, DOI: 10.1007/s10959-022-01234-5, In press.