A SHORT NOTE ON ASYMPTOTIC ENUMERATION OF CONTINGENCY TABLES WITH NON-UNIFORM MARGINS

DA WU

Abstract. In this short note, we compute the precise asymptotics for the number of contingency tables with non-uniform margins. More precisely, for parameter \( n, \delta, B, C > 0 \), we consider the set of matrices whose first \( \lfloor n^\delta \rfloor \) rows and columns have sum \( BCn \) and the remaining \( n \) rows and columns have sum \( Cn \). We compute the precise asymptotics of the cardinality of this set when \( B < B_c = 1 + \sqrt{1 + 1/C} \) using the maximal entropy principle introduced in [1]. The only contribution of this note is a detailed expansion of the determinant of quadratic forms in asymptotic formulas.

1. Introduction

Let \( \mathbf{r} = (r_1, \ldots, r_m) \) and \( \mathbf{c} = (c_1, \ldots, c_n) \) be two positive integer vectors such that
\[
\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = N.
\]
Let \( M(\mathbf{r}, \mathbf{c}) \) be the set of \( m \times n \) non-negative integer matrices with \( i \)-th row sum \( r_i \) and \( j \)-th column sum \( c_j \) for all \( 1 \leq i \leq m, 1 \leq j \leq n \). Suppose all the \( r_i \) and \( c_j \) depend on the dimension \( m \) and \( n \), one of the fundamental problems in Combinatorics is to provide the precise asymptotics of \( \#M(\mathbf{r}, \mathbf{c}) \) as \( m, n \to \infty \). Recently, following the work by Pak and Lyu in [2] we are interested in the case of non-uniform margin with two different values. More precisely, we consider the case when
\[
\tilde{\mathbf{r}} = \tilde{\mathbf{c}} = (\lfloor BCn \rfloor, \ldots, \lfloor BCn \rfloor, \lfloor Cn \rfloor, \ldots, \lfloor Cn \rfloor) \in \mathbb{N}^{\lfloor n^\delta \rfloor + n}
\]
for parameters \( B, C > 0 \) and \( 0 \leq \delta < 1 \). Let
\[
M_{n,\delta}(B, C) := M(\tilde{\mathbf{r}}, \tilde{\mathbf{c}}).
\]
We are interested in precise asymptotics of \( \#M_{n,\delta}(B, C) \) when \( n \to \infty \). It is shown in [2] that the typical table (defined in [1]) associated with \( \tilde{\mathbf{r}} \) and \( \tilde{\mathbf{c}} \) are uniform bounded in large \( n \) limit when \( B < B_c = 1 + \sqrt{1 + 1/C} \). In this case, we apply the maximal entropy method in [1] to compute the precise asymptotics of \( M_{n,\delta}(B, C) \). When \( B > B_c \), since entries in top left corner will blow up in large \( n \) limit, the precise asymptotics is not known. However, loose estimate of \( \log \#M_{n,\delta}(B, C) \) is known with error \( O(n \log n + n^{2\delta}) \), see the main theorem in [3].

2. Precise Asymptotics of \( \#M_{n,\delta}(B, C) \) in Sub-critical Regime

In this section, we compute the precise asymptotic formula for \( \#M_{n,\delta}(B, C) \) when \( 0 \leq \delta < 1 \) and \( B < B_c = 1 + \sqrt{1 + 1/C} \) (subcritical case). The computation is based on Theorem 1.3 in [1] and Lemma 5.1 in [2], which will be restated below.
2.1. Review of Literature. First, we recall the general asymptotic formula for \( \#M(r, c) \) when all of the entries of typical table \( Z \) are of same order. More detailed description can be found in [1] Section 1. We say margins \( r = (r_1, \ldots, r_m) \) and \( c = (c_1, \ldots, c_n) \) are \( \delta' \text{-smooth} \) if they satisfy the following two conditions:

(i): \( m \geq \delta'n \) and \( n \geq \delta'm \). Namely, dimensions of the matrix are of the same order asymptotically.

(ii): \( \delta'\tau \leq z_{ij} \leq \tau \) for some \( \tau \) such that \( \tau \geq \delta' \) and all \( 1 \leq i \leq m, 1 \leq j \leq n \). Namely, entries of typical table are of the same order asymptotically.

Next, for typical table \( Z = (z_{ij}) \) associated with \( M(r, c) \), we define the quadratic form \( q : \mathbb{R}^{m+n} \to \mathbb{R} \) as the following:

\[
q(s, t) := \frac{1}{2} \sum_{1 \leq j \leq m} \sum_{1 \leq k \leq n} (z_{jk}^2 + z_{jk})(s_j + t_k)^2,
\]

where \( s = (s_1, \ldots, s_m) \) and \( t = (t_1, \ldots, t_n) \). Notice that the null space is spanned by the vector \( \mathbf{v} = (1, \ldots, 1, -1, \ldots, -1) \). Let \( H = u^\perp \subseteq \mathbb{R}^{m+n} \) and \( q|_H \) is a positive definite quadratic form and we can define its determinant \( \det(q|_H) \) to be the product of non-zero eigenvalues of \( q \).

We also define the polynomials \( f, h : \mathbb{R}^{m+n} \to \mathbb{R} \) by

\[
f(s, t) := \frac{1}{6} \sum_{1 \leq j \leq m} \sum_{1 \leq k \leq n} z_{jk}(z_{jk} + 1)(2z_{jk} + 1)(s_j + t_k)^3
\]

and

\[
h(s, t) := \frac{1}{24} \sum_{1 \leq j \leq m} \sum_{1 \leq k \leq n} z_{jk}(z_{jk} + 1)(6z_{jk}^2 + 6z_{jk} + 1)(s_j + t_k)^4,
\]

where \( s = (s_1, \ldots, s_m) \) and \( t = (t_1, \ldots, t_n) \). Consider the Gaussian probability measure on \( H \) with density proportional to \( e^{-q} \) and define

\[
\mu := \mathbb{E}[f^2] \quad \text{and} \quad \nu := \mathbb{E}[h].
\]

Now, we can state the main theorem in [1].

**Theorem 2.1** ([1], Theorem 1.3). Fix \( 0 < \delta' < 1 \) and let \( r \) and \( c \) be \( \delta' \text{-smooth} \) margins and \( Z = (z_{ij}) \) be the associated typical table for \( M(r, c) \). Then

\[
\#M(r, c) \asymp \frac{e^{q(Z)} \sqrt{m+n}}{(4\pi)^{(m+n-1)/2} \sqrt{\det(q|_H)}} \exp\left(-\frac{\mu}{2} + \nu\right)
\]
as \( m, n \to +\infty \).

**Remark 2.2.** There exists some positive constants \( \gamma_1(\delta') \) and \( \gamma_2(\delta') \) such that

\[
\gamma_1(\delta') \leq \exp\left(-\frac{\mu}{2} + \nu\right) \leq \gamma_2(\delta').
\]

Therefore,

\[
\exp\left(-\frac{\mu}{2} + \nu\right) = O(1).
\]
Remark 2.3. Using the change of coordinate basis,
\begin{equation}
\text{det}(q|_M) = (m + n) \cdot 2^{1-m-n} \text{det} Q,
\end{equation}
where $Q = (q_{ij})$ is the $(m + n - 1) \times (m + n - 1)$ symmetric matrix with
\begin{equation}
q_{j,k+m} = q_{k+m,j} = z_{jk}^2 + z_{jk} \quad \text{for } 1 \leq j \leq m, 1 \leq k \leq n - 1,
\end{equation}
\begin{equation}
q_{jj} = r_j + \sum_{k=1}^{n} z_{jk}^2 = \sum_{k=1}^{n} (z_{jk} + z_{jk}^2) \quad \text{for } 1 \leq j \leq m,
\end{equation}
\begin{equation}
q_{k+m,k+m} = c_k + \sum_{j=1}^{n} z_{jk}^2 = \sum_{j=1}^{n} (z_{jk} + z_{jk}^2) \quad \text{for } 1 \leq k \leq n - 1.
\end{equation}
Therefore, we can further simplify (2.3) to
\begin{equation}
\#M(r, c) \propto \frac{e^{g(Z)}}{(2\pi)^{(m+n-1)/2} \sqrt{\text{det} Q}} \exp \left( -\frac{\mu}{2} + \nu \right).
\end{equation}
See [1] Section 1.4 for a more detailed explanation.

Next, we recall the key Lemma in [2] regarding the asymptotics of entries of $Z = (z_{ij})$ associated with $M_{n,\delta}(B,C)$.

Lemma 2.4 ([2], Lemma 5.1). Fix $0 \leq \delta < 1$ and let $Z = (z_{ij})_{1 \leq i,j \leq n + [n^\delta]}$ be the typical table of $M_{n,\delta}(B,C)$. Let $B_c = 1 + \sqrt{1 + 1/C}$ and we have the following,
(i): If $B < B_c$, then
\begin{equation}
z_{11} = \frac{B^2(C + 1)}{(B_c - B)(B_c + B - 2)} + O(n^{\delta-1}), \quad z_{1,n+1} = BC + O(n^{\delta-1}).
\end{equation}
(ii): If $B > B_c$, then
\begin{equation}
z_{n+1,n+1} = C + O(n^{\delta-1}), \quad z_{1,n+1} = B_cC + O(n^{\delta-1}), \quad n^{\delta-1}z_{11} = C(B - B_c) + O(n^{\delta-1}).
\end{equation}

Remark 2.5. The behaviour of $z_{n+1,n+1}$ is more predictable. It is shown in [1] that
\begin{equation}
|z_{n+1,n+1} - C| = n^{\delta-1}z_{1,n+1} \leq BCn^{\delta-1}.
\end{equation}
Hence, it is trivial that
\begin{equation}
z_{n+1,n+1} = C + O(n^{\delta-1}).
\end{equation}

2.2. Computation of $\#M_{n,\delta}(B,C)$. Now, we go back to our setting of $\#M_{n,\delta}(B,C)$. Recall $0 \leq \delta < 1$ and $B < B_c = 1 + \sqrt{1 + 1/C}$. First, notice that when $B < B_c$, all of entries of $Z = (z_{ij})$ have well-defined finite limits, and by symmetry
\begin{equation}
e^{g(Z)} = \prod_{1 \leq i,j \leq n + [n^\delta]} \frac{(z_{ij} + 1)^{z_{ij}+1}}{z_{ij}^{z_{ij}}}
= \left( \frac{z_{11} + 1}{z_{11}^{z_{11}}} \right)^{[n^\delta]^2} \left( \frac{(z_{n+1,n+1} + 1)^{z_{n+1,n+1}+1}}{z_{n+1,n+1}^{z_{n+1,n+1}} \cdot z_{n+1,n+1}^{z_{n+1,n+1}}} \right)^n \left( \frac{(z_{1,n+1} + 1)^{z_{1,n+1}+1}}{z_{1,n+1}^{z_{1,n+1}} \cdot z_{1,n+1}^{z_{1,n+1}}} \right)^{2n[n^\delta]}.
\end{equation}
Next, we compute the determinant of $Q$ in (2.7). By (2.6), $Q$ has entries

\begin{equation}
q_{jj} = [BCn] + [n^\delta] \left( \frac{B^2(C + 1)}{(B_c - B)(B_c + B - 2)} + O(n^{\delta-1}) \right)^2 + n (BC + O(n^{\delta-1}))^2
\end{equation}

when $1 \leq j \leq [n^\delta]$ and $[n^\delta] + n + 1 \leq j \leq 2[n^\delta] + n$

\begin{equation}
q_{jj} = [Cn] + [n^\delta] (BC + O(n^{\delta-1}))^2 + n(C + O(n^{\delta-1}))^2
\end{equation}

when $[n^\delta] + 1 \leq j \leq [n^\delta] + n$ and $2[n^\delta] + n + 1 \leq j \leq 2([n^\delta] + n) - 1$.

\begin{equation}
q_{ij} = q_{ji} = \left( \frac{B^2(1 + C)}{(B_c - B)(B_c + B - 2)} + O(n^{\delta-1}) \right)^2 + \frac{B^2(1 + C)}{(B_c - B)(B_c + B - 2)} + O(n^{\delta-1})
\end{equation}

when $1 \leq i \leq [n^\delta]$ and $[n^\delta] + n + 1 \leq j \leq 2[n^\delta] + n$.

\begin{equation}
q_{ij} = q_{ji} = (BC + O(n^{\delta-1}))^2 + BC + O(n^{\delta-1})
\end{equation}

when $1 \leq i \leq [n^\delta]$, $2[n^\delta] + n + 1 \leq j \leq 2([n^\delta] + n) - 1$ and when $[n^\delta] + 1 \leq i \leq [n^\delta] + n$, $[n^\delta] + n + 1 \leq j \leq 2[n^\delta] + n$.

\begin{equation}
q_{ij} = q_{ji} = (C + O(n^{\delta-1}))^2 + C + O(n^{\delta-1})
\end{equation}

when $[n^\delta] + 1 \leq i \leq [n^\delta] + n$ and $2[n^\delta] + n + 1 \leq j \leq 2([n^\delta] + n) - 1$.

The rest of the entries are zero. Notice that all the off-diagonal entries have size $O(1)$ while all the entries on the diagonal has asymptotical order $n$. To compute the asymptotics of $\det Q$, we write $Q = A + E$ where $A = \text{diag}(q_{11}, q_{22}, \ldots, q_{2([n^\delta] + n) - 1, 2([n^\delta] + n) - 1})$ is the diagonal matrix. By diagonal expansion of the determinant,

\begin{equation}
\det(Q) = \det(A+E) = \det(A) + S_1 + S_2 + \ldots + S_{2([n^\delta] + n - 1)} + \det(E)
\end{equation}

where

\[ S_k = \sum_{1 \leq i_1 < \ldots < i_k \leq 2([n^\delta] + n) - 1} \left( \prod_{r=1}^{k} q_{i_r, i_r} \right) \det(E_{i_1, \ldots, i_k}) \]

$E_{i_1, \ldots, i_k}$ is the principle minor of order $2([n^\delta] + n) - 1 - k$ of $E$. Trivially, $\det(E) = 0$, and

\begin{equation}
\det(A) = \prod_{i=1}^{2([n^\delta] + n) - 1} q_{ii}
\end{equation}
Furthermore, \( S_{2([n^δ]+n)-3} = 0 \), and
\[
S_{2([n^δ]+n)-4} = 0
\]
\[
S_{2([n^δ]+n)-5} = ([n^δ])^2 n(n-1) \left[(z_{11}^2 + z_{11})(z_{n+1,n+1}^2 + z_{n+1,n+1}) - (z_{1,n+1}^2 + z_{1,n+1})^2\right]
\]
\[
S_{2([n^δ]+n)-i} = 0 \quad \forall i \geq 6
\]
The above computation is based on the symmetry of typical table, i.e.
\[
z_{ij} = z_{i'j'} \quad \text{if } r_i = r_{i'} \text{ and } c_j = c_{j'}
\]
Therefore,
\[
(2.20) \quad \det(Q) = \left(\prod_{i=1}^{2([n^δ]+n)-1} q_i\right) - (q_1 \cdots q_{[n^δ]+n-1}) (q_{[n^δ]+n+2} \cdots q_{2[n^δ]+2n-1}) \left[(z_{1,n+1}^2 + z_{1,n+1})^2\right]
\]
\[
+ ([n^δ])^2 n(n-1) \left[(z_{11}^2 + z_{11})(z_{n+1,n+1}^2 + z_{n+1,n+1}) - (z_{1,n+1}^2 + z_{1,n+1})^2\right]
\]
where we write \( q_i \) in place of \( q_{ii} \). Finally, by (2.13), (2.14), (2.15), (2.16), (2.17), (2.7) and Lemma 2.4 we get the precise asymptotics of \( \#M_{n,δ}(B,C) \).

### 3. Left half of the \( M_{n,δ}(B,C) \)

In this section, we compute the case when
\[
r_1 = \begin{pmatrix} B_n C_n, \ldots, B_n C_n \\ \frac{[n^δ]}{n^δ} \text{ entries} \\ C_n - B_n C_n \delta, \ldots, C_n - B_n C_n \delta \end{pmatrix} \in \mathbb{Z}_{>0}^{n^δ+n^δ}
\]
and
\[
c_1 = (C_n, \ldots, C_n) \in \mathbb{Z}_{>0}^n
\]
By symmetry and margin conditions, \( Z = (z_{ij}) \) satisfies
\[
\left\{ \\
\begin{array}{c}
n z_{11} = B_n C_n \\
n^δ z_{11} + n z_{1,n+1} = C n
\end{array} \right.
\]
which implies that
\[
\left\{ \\
\begin{array}{c}
z_{11} = B_n C \\
z_{1,n+1} = C - z_{11} n^{δ-1} = C - B_n C n^{δ-1}
\end{array} \right.
\]
Next, we compute the exact asymptotic formula of \( \#M(r_1, c_1) \). Recall the formula,
\[
\frac{e^{θ(Z)}}{(2π)^{(m+n-1)/2} \sqrt{det(Q)}} \exp \left(-\frac{μ}{2} + ν\right)
\]
where \( Q = (q_{ij}) \in \mathbb{R}^{(2n+n^\delta-1) \times (2n+n^\delta-1)} \) has entries

\[
q_{ii} = \begin{cases} 
B_c C n + n (B_c C)^2 & 1 \leq i \leq [n^\delta] \\
C n - B_c C n^\delta + n (C - B_c C n^\delta - 1)^2 & [n^\delta] + 1 \leq i \leq [n^\delta] + n \\
C n + n^\delta (B_c C)^2 + (n - n^\delta) (C - B_c C n^\delta - 1)^2 & [n^\delta] + n + 1 \leq i \leq 2n + [n^\delta] - 1 
\end{cases}
\]

and

\[
q_{ij} = q_{ji} = \begin{cases} 
B_c C + (B_c C)^2 & 1 \leq i \leq [n^\delta], n + [n^\delta] + 1 \leq j \leq 2n + [n^\delta] - 1 \\
C - B_c C n^\delta - 1 + (C - B_c C n^\delta - 1)^2 & [n^\delta] + 1 \leq i \leq [n^\delta] + n, n + [n^\delta] + 1 \leq j \leq 2n + [n^\delta] - 1 
\end{cases}
\]

The rest of the entries are 0. We write \( Q = A + E \) where

\[
A = \text{diag} (q_{11}, \ldots, q_{2n+[n^\delta]-1,2n+[n^\delta]-1})
\]

By diagonal expansion of the determinants,

\[
det(Q) = det(A + E) = det(A) + S_1 + S_2 + \ldots + S_{2n+[n^\delta]-2} + det(E)
\]

where

\[
S_k = \sum_{1 \leq i_1 < \ldots < i_k \leq 2n+[n^\delta]-2} \left( \prod_{r=1}^{k} q_{i_r,i_r} \right) \det (E_{i_1,\ldots,i_k})
\]

\( E_{i_1,\ldots,i_k} \) is the principle minor of order \( 2n + [n^\delta] - 2 - k \) of \( E \). It is not hard to see that

\[
S_{2n+[n^\delta]-2} = 0 \\
S_{2n+[n^\delta]-3} \\
\quad = -[n^\delta](n-1) \left\{ B_c C n + n (B_c C)^2 \right\}^{[n^\delta]-1} \left\{ C n - B_c C n^\delta + n (C - B_c C n^\delta - 1)^2 \right\} \times \\
\quad \left\{ C n + n^\delta (B_c C)^2 + (n - n^\delta) (C - B_c C n^\delta - 1)^2 \right\}^{n-2} (B_c C + (B_c C)^2)^2 \\
\quad - n(n-1) \left\{ B_c C n + n (B_c C)^2 \right\}^{[n^\delta]} \left\{ C n - B_c C n^\delta + n (C - B_c C n^\delta - 1)^2 \right\}^{n-1} \times \\
\quad \left\{ C n + n^\delta (B_c C)^2 + (n - n^\delta) (C - B_c C n^\delta - 1)^2 \right\}^{n-2} \left( C - B_c C n^\delta - 1 + (C - B_c C n^\delta - 1)^2 \right)^2
\]

\( S_{2n+[n^\delta]-i} = 0 \)
for all \( i \geq 4 \). Therefore,
\[
\det Q = \det A + S_{2n+[n^\delta]-3}
\]
\[
= (q_{11})^{[n^\delta]} \left( q_{n+1,n+1} \right)^n (q_{2n+1,2n+1})^{n-1}
\]
\[
- [n^\delta](n-1)(q_{11})^{[n^\delta]-1} \left( q_{n+1,n+1} \right)^n (q_{2n+1,2n+1})^{n-2} \left( B_c C + (B_c C)^2 \right)^2
\]
\[
- n(n-1)(q_{11})^{[n^\delta]} \left( q_{n+1,n+1} \right)^n (q_{2n+1,2n+1})^{n-2} \left( C - B_c C n^{\delta-1} + (C - B_c C n^{\delta-1})^2 \right)^2
\]
\[
= \{ B_c C n + n (B_c C)^2 \}^{[n^\delta]} \left\{ Cn - B_c C n^{\delta} + n \left( C - B_c C n^{\delta-1} \right)^2 \right\} \times \n
\]
\[
\left\{ Cn + n^\delta (B_c C)^2 + (n - n^\delta) \left( C - B_c C n^{\delta-1} \right)^2 \right\}^{n-1}
\]
\[
- [n^\delta](n-1) \left\{ B_c C n + n (B_c C)^2 \right\}^{[n^\delta]-1} \left\{ Cn - B_c C n^{\delta} + n \left( C - B_c C n^{\delta-1} \right)^2 \right\} \times \n
\]
\[
\left\{ Cn + n^\delta (B_c C)^2 + (n - n^\delta) \left( C - B_c C n^{\delta-1} \right)^2 \right\}^{n-2} \left( B_c C + (B_c C)^2 \right)^2
\]
\[
- n(n-1) \left\{ B_c C n + n (B_c C)^2 \right\}^{[n^\delta]} \left\{ Cn - B_c C n^{\delta} + n \left( C - B_c C n^{\delta-1} \right)^2 \right\}^{n-1} \times \n
\]
\[
\left\{ Cn + n^\delta (B_c C)^2 + (n - n^\delta) \left( C - B_c C n^{\delta-1} \right)^2 \right\}^{n-2} \left( C - B_c C n^{\delta-1} + (C - B_c C n^{\delta-1})^2 \right)^2
\].

Plugging in (2.7) and we are done.

REFERENCES

[1] A. Barvinok and J.A. Hartigan, An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums, Trans. AMS, Vol. 364 (2012), 4323-4368

[2] S. Dittmer, H. Lyu and I. Pak, Phase transition in random contingency tables with non-uniform margins, to appear in Trans. AMS; arXiv: 1903. 08743.

[3] H. Lyu and I. Pak, On the number of contingency tables and the independent heuristic, arXiv:2009.10810

University of Pennsylvania, Department of Mathematics, David Rittenhouse Lab, 209 South 33rd Street, Philadelphia, PA, 10104-6395

Email address: dawu@math.upenn.edu