On the number of contingency tables and the independence heuristic

Hanbaek Lyu

University of Wisconsin - Madison

Joint work with Igor Pak and Sumit Mukherjee

July 24, 2023

Outline

Introduction

Indpendence heuristic and second-order phase transition

Barvinok's conjecture and first-order phase transition

Typical table

Strong duality between typical table and MLE

Contingency tables

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.

Contingency tables

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.
- margins: $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{N}^m$, $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{N}^n$, $\sum a_i = \sum b_j = N$

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.
- margins: $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{N}^m$, $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{N}^n$, $\sum a_i = \sum b_j = N$
- Let $\mathcal{T}(\mathbf{a}, \mathbf{b})$ be the set of all $(n \times n)$ contingency tables of row sum $\mathbf{a}$ and column sum $\mathbf{b}$:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) := \left\{ (x_{ij}) \in \mathbb{N}^{n \times n} \,\middle|\, \sum_{k=1}^{n} x_{ik} = a_i, \sum_{k=1}^{n} x_{kj} = b_j \quad \forall 1 \leq i, j \leq n \right\}$$

Contingency tables

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.
- margins: $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{N}^m$, $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{N}^n$, $\sum a_i = \sum b_j = N$
- Let $\mathcal{T}(\mathbf{a}, \mathbf{b})$ be the set of all $(n \times n)$ contingency tables of row sum $\mathbf{a}$ and column sum $\mathbf{b}$:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) := \left\{ (x_{ij}) \in \mathbb{N}^{n \times n} \,\middle|\, \sum_{k=1}^{n} x_{ik} = a_i, \sum_{k=1}^{n} x_{kj} = b_j \quad \forall 1 \leq i, j \leq n \right\}$$

  - $\mathrm{T}(\mathbf{a}, \mathbf{b}) := |\mathcal{T}(\mathbf{a}, \mathbf{b})|$ $(= \#$ of bipartite graphs with degree sequences $\mathbf{a}$ and $\mathbf{b})$

Contingency tables

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.
- margins: $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{N}^m$, $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{N}^n$, $\sum a_i = \sum b_j = N$
- Let $\mathcal{T}(\mathbf{a}, \mathbf{b})$ be the set of all $(n \times n)$ contingency tables of row sum $\mathbf{a}$ and column sum $\mathbf{b}$:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) := \left\{ (x_{ij}) \in \mathbb{N}^{n \times n} \,\middle|\, \sum_{k=1}^{n} x_{ik} = a_i, \sum_{k=1}^{n} x_{kj} = b_j \quad \forall 1 \leq i, j \leq n \right\}$$

  - $\mathrm{T}(\mathbf{a}, \mathbf{b}) := |\mathcal{T}(\mathbf{a}, \mathbf{b})|$ ($=$ # of bipartite graphs with degree sequences $\mathbf{a}$ and $\mathbf{b}$)
- Sampling CT: $X \sim \text{Uniform}(\mathcal{T}(\mathbf{a}, \mathbf{b}))$

Contingency tables

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.
- margins: $\mathbf{a} = (a_1, \dots, a_m) \in \mathbb{N}^m$, $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{N}^n$, $\sum a_i = \sum b_j = N$
- Let $\mathcal{T}(\mathbf{a}, \mathbf{b})$ be the set of all $(n \times n)$ contingency tables of row sum $\mathbf{a}$ and column sum $\mathbf{b}$:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) := \left\{ (x_{ij}) \in \mathbb{N}^{n \times n} \,\middle|\, \sum_{k=1}^{n} x_{ik} = a_i, \sum_{k=1}^{n} x_{kj} = b_j \quad \forall 1 \leq i, j \leq n \right\}$$

  - $\mathrm{T}(\mathbf{a}, \mathbf{b}) := |\mathcal{T}(\mathbf{a}, \mathbf{b})|$ ($=$ # of bipartite graphs with degree sequences $\mathbf{a}$ and $\mathbf{b}$)
- Sampling CT: $X \sim \text{Uniform}(\mathcal{T}(\mathbf{a}, \mathbf{b}))$
- Counting CT: Compute $\mathrm{T}(\mathbf{a}, \mathbf{b})$

- Contingency tables are matrices with non-netative integer entries with fixed row an column margins.
- margins: $\mathbf{a} = (a_1, \dots, a_m) \in \mathbb{N}^m$, $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{N}^n$, $\sum a_i = \sum b_j = N$
- Let $\mathcal{T}(\mathbf{a}, \mathbf{b})$ be the set of all $(n \times n)$ contingency tables of row sum $\mathbf{a}$ and column sum $\mathbf{b}$:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) := \left\{ (x_{ij}) \in \mathbb{N}^{n \times n} \,\middle|\, \sum_{k=1}^{n} x_{ik} = a_i, \sum_{k=1}^{n} x_{kj} = b_j \quad \forall 1 \leq i, j \leq n \right\}$$

  - $\mathrm{T}(\mathbf{a}, \mathbf{b}) := |\mathcal{T}(\mathbf{a}, \mathbf{b})|$ ($=$ $\#$ of bipartite graphs with degree sequences $\mathbf{a}$ and $\mathbf{b}$)
- Sampling CT: $X \sim \text{Uniform}(\mathcal{T}(\mathbf{a}, \mathbf{b}))$
- Counting CT: Compute $\mathrm{T}(\mathbf{a}, \mathbf{b})$
- Sampling $\leftrightarrow$ Counting (self-reduction):

$$\mathbb{P}\left(X_{11} \geq t\right) = \frac{\mathrm{T}\begin{pmatrix} \mathbf{a} = (a_1 - t, a_2 \dots, a_m) \\ \mathbf{b} = (b_1 - t, b_2 \dots, b_n) \end{pmatrix}}{\mathrm{T}\begin{pmatrix} \mathbf{a} = (a_1, a_2 \dots, a_m) \\ \mathbf{b} = (b_1, b_2 \dots, b_n) \end{pmatrix}}$$

Data

1	0	3	2	0	7	13
1	2	0	4	3	0	10
7	5	2	1	0	0	15
0	0	3	1	3	9	16
0	3	1	8	0	2	14
5	3	0	3	5	3	19
9	13	9	19	11	21	

$v.s.$

Null model

						13
						10
						15
		$X = (X_{ij})$				16
						14
						19
9	13	9	19	11	21	

- Contingency tables are fundamental tools in statistics for studying dependence structure between two or more variables

Uniform contingency tables in statistics

Data

1	0	3	2	0	7	13
1	2	0	4	3	0	10
7	5	2	1	0	0	15
0	0	3	1	3	9	16
0	3	1	8	0	2	14
5	3	0	3	5	3	19
9	13	9	19	11	21	

$v.s.$

Null model

						13
						10
						15
		$X = (X_{ij})$				16
						14
						19
9	13	9	19	11	21	

- Contingency tables are fundamental tools in statistics for studying dependence structure between two or more variables
- Uniform contingency table $X = (X_{ij})$ serves as the maximum entropy null model given margins

Data

1	0	3	2	0	7	13
1	2	0	4	3	0	10
7	5	2	1	0	0	15
0	0	3	1	3	9	16
0	3	1	8	0	2	14
5	3	0	3	5	3	19
9	13	9	19	11	21	

$v.s.$

Null model

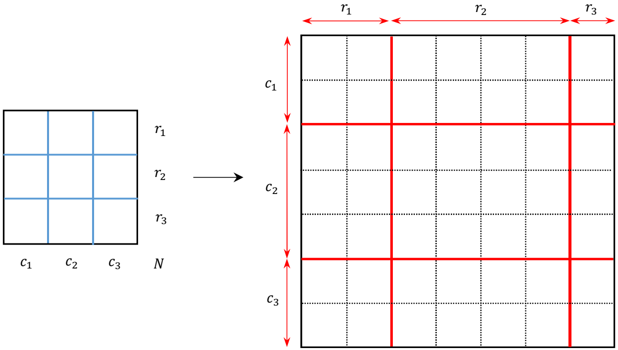
						13
						10
						15
						16
						14
						19
9	13	9	19	11	21	

$X = (X_{ij})$

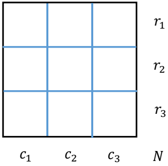
- Contingency tables are fundamental tools in statistics for studying dependence structure between two or more variables
- Uniform contingency table $X = (X_{ij})$ serves as the maximum entropy null model given margins
- It motivates to study the structure of X for given margins

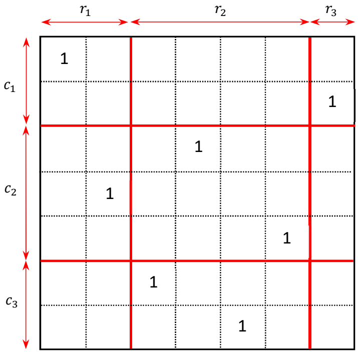
Sampling random CTs in statistics

- To sample from $\mathcal{T}(\mathbf{a}, \mathbf{b})$, first consider a 0-1 block matrix of size $N \times N = (a_1 + \cdots + a_m) \times (b_1 + \cdots + b_n)$:

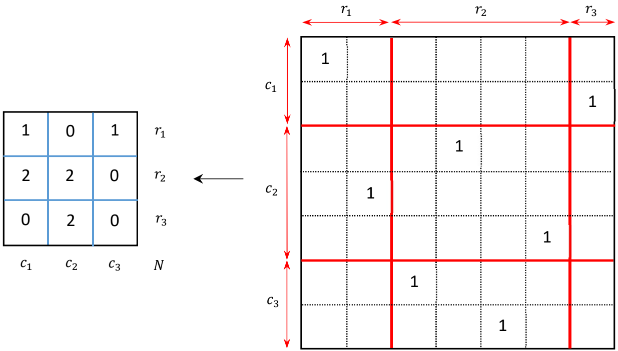


- Fill in the block matrix with a uniform random permutation matrix:





- Collapse each block into each cell in the contingency table



- Resulting contingency table follows hypergeometric distribution: (not uniform)

$$\mathbb{P}(Y = (y_{ij})) \propto \prod_{ij} \frac{1}{y_{ij}!}$$

Counting CTs — Numerical examples (Uniform margins)

- row sums $= s$, column sums $= t$, total sum $= ms = nt\,(= N)$

Case	m	n	s	t	UB1	UB2	UB3	**Actual**	**New LB**	LB2	LB1
1	3	3	100	100	4.7×10^{17}	1.8×10^{15}	3.4×10^{11}	1.3×10^{7}	3.1×10^{5}	2.4×10^{3}	1.5×10^{-28}
2	3	9	99	33	2.3×10^{40}	1.5×10^{38}	3.7×10^{29}	2.8×10^{21}	7.3×10^{17}	5.6×10^{15}	1.2×10^{-62}
3	3	49	98	6	8.1×10^{121}	1.1×10^{120}	1.1×10^{98}	1.0×10^{68}	9.1×10^{55}	6.4×10^{53}	4.1×10^{-381}
4	10	10	20	20	8.5×10^{82}	1.4×10^{81}	2.2×10^{74}	1.1×10^{59}	5.7×10^{49}	4.8×10^{41}	5.2×10^{-104}
5	18	18	13	13	6.4×10^{164}	1.3×10^{163}	6.0×10^{156}	7.9×10^{127}	1.1×10^{110}	2.7×10^{95}	1.1×10^{-214}
6	30	30	3	3	9.5×10^{130}	3.8×10^{129}	3.8×10^{128}	2.2×10^{92}	2.2×10^{73}	1.6×10^{56}	2.2×10^{-522}
7	100	100	3	3	1.2×10^{589}	2.8×10^{587}	3.4×10^{586}	5.3×10^{459}	4.9×10^{394}	4.1×10^{332}	1.5×10^{-2267}
8	4	4	300	300	9.9×10^{36}	1.3×10^{34}	5.1×10^{25}	2.0×10^{19}	4.1×10^{16}	3.8×10^{12}	2.5×10^{-39}
9	9	9	10^3	10^3	1.1×10^{201}	4.4×10^{197}	1.8×10^{168}	8.0×10^{151}	4.5×10^{142}	7.3×10^{128}	1.8×10^{-32}
10	9	9	10^5	10^5	7.7×10^{362}	3.1×10^{357}	1.4×10^{298}	6.1×10^{279}	3.2×10^{270}	5.2×10^{248}	1.5×10^{44}
11	15	15	10^3	10^3	6.7×10^{508}	2.6×10^{505}	3.8×10^{457}	$\approx 1.7 \times 10^{427}$	1.7×10^{409}	2.3×10^{384}	1.3×10^{80}
12	15	15	10^5	10^5	1.3×10^{958}	5.1×10^{952}	1.1×10^{851}	$\approx 1.7 \times 10^{819}$	3.2×10^{800}	4.5×10^{761}	4.0×10^{383}
13	100	100	10^3	10^3	1.3×10^{14553}	6.0×10^{14549}	8.2×10^{14346}	$\approx 6.3 \times 10^{14072}$	5.3×10^{13869}	4.6×10^{13684}	5.0×10^{10741}
14	100	100	10^5	10^5	1.3×10^{34345}	5.2×10^{34339}	1.1×10^{33751}	$\approx 6.3 \times 10^{33470}$	4.9×10^{33263}	4.4×10^{32979}	6.2×10^{29545}

Figure: Excerpted from [3]

- UB1, LB1 = Barvinok's first upper and lower bounds [1]
- UB2, LB2 = Barvinok's first upper and lower bounds [2]
- UB3 = Shapiro's upper bound [13]
- New LB = Brändén, Leake, Pak [3]

Counting TCs — Numerical examples (Non-uniform margins)

Case	m	n	N	UB1	UB2	UB3	**Actual**	**New LB**	LB2	LB1	Time
1	4	4	592	3.0×10^{30}	6.0×10^{27}	7.1×10^{18}	1.2×10^{15}	9.5×10^{12}	4.6×10^{8}	3.8×10^{-40}	79 sec
2	5	4	1269	1.4×10^{34}	1.2×10^{31}	8.3×10^{20}	3.4×10^{16}	2.0×10^{14}	3.0×10^{7}	1.5×10^{-52}	550 sec
3	4	4	65159458	1.3×10^{112}	?	2.1×10^{65}	4.3×10^{61}	5.8×10^{58}	?	2.3×10^{-49}	N/A
4	50	50	486	7.2×10^{562}	?	1.3×10^{551}	??	5.2×10^{421}	?	6.4×10^{-749}	N/A
5	50	50	302	1.2×10^{350}	?	7.3×10^{338}	??	1.1×10^{239}	?	2.0×10^{-922}	N/A

Figure: Excerpted from [3]

- UB1, LB1 = Barvinok's first upper and lower bounds [1]
- UB2, LB2 = Barvinok's first upper and lower bounds [2]
- UB3 = Shapiro's upper bound [13]
- New LB = Brändén, Leake, Pak [3]

- Large gap between rigorous upper and lower bounds on $\mathrm{T}(\mathbf{a}, \mathbf{b})$ for non-uniform margins

Uniform and smooth margins

Uniform margins: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.

Uniform and smooth margins

Uniform margins: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.

- Sharp volume estimate (Canfield and MacKay '10 [4]):

$$\log \mathrm{T}(\mathbf{a}, \mathbf{b}) = [(1+C)\log(1+C) - C\log(C)]n^2 - n\log n \\ - n\log 2\pi C(1+C) + \log n + O(1).$$

Uniform and smooth margins

Uniform margins: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.

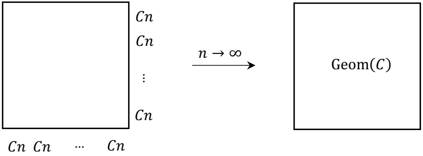
- Sharp volume estimate (Canfield and MacKay '10 [4]):

$$\log \mathrm{T}(\mathbf{a}, \mathbf{b}) = [(1+C)\log(1+C) - C\log(C)]n^2 - n\log n \\ - n\log 2\pi C(1+C) + \log n + O(1).$$

- Convergence to geometric RVs of mean C (Chatterjee, Diaconis, and Sly '10 [5]):

$$d_{TV}(X_{ij}, \mathrm{Geom}(C)) \to 0 \quad \text{as } n \to \infty$$

Asymptotically independent entries



Uniform margins: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.

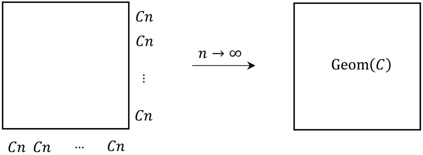
- Sharp volume estimate (Canfield and MacKay '10 [4]):

$$\log \mathrm{T}(\mathbf{a}, \mathbf{b}) = [(1+C)\log(1+C) - C\log(C)]n^2 - n\log n \\ - n\log 2\pi C(1+C) + \log n + O(1).$$

- Convergence to geometric RVs of mean C (Chatterjee, Diaconis, and Sly '10 [5]):

$$d_{TV}(X_{ij}, \mathrm{Geom}(C)) \to 0 \quad \text{as } n \to \infty$$

Asymptotically independent entries



- Empirical distribution of eigenvalues $\Rightarrow$ circular law (Nguyen '14 [12])

Uniform and smooth margins

Uniform margins: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.

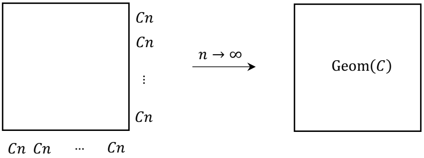
- Sharp volume estimate (Canfield and MacKay '10 [4]):

$$\log \mathrm{T}(\mathbf{a}, \mathbf{b}) = [(1+C)\log(1+C) - C\log(C)]n^2 - n\log n - n\log 2\pi C(1+C) + \log n + O(1).$$

- Convergence to geometric RVs of mean C (Chatterjee, Diaconis, and Sly '10 [5]):

$$d_{TV}(X_{ij}, \mathrm{Geom}(C)) \to 0 \quad \text{as } n \to \infty$$

Asymptotically independent entries



- Empirical distribution of eigenvalues $\Rightarrow$ circular law (Nguyen '14 [12])

Smooth margins: $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$ so that $\frac{\max \mathbf{a}}{\min \mathbf{a}}, \frac{\max \mathbf{b}}{\min \mathbf{b}} \leq \phi = (1+\sqrt{5})/2 \approx 1.618$.

Uniform and smooth margins

Uniform margins: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.

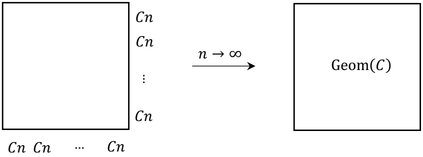
- Sharp volume estimate (Canfield and MacKay '10 [4]):

$$\log \mathrm{T}(\mathbf{a}, \mathbf{b}) = [(1+C)\log(1+C) - C\log(C)]n^2 - n\log n$$
$$- n\log 2\pi C(1+C) + \log n + O(1).$$

- Convergence to geometric RVs of mean C (Chatterjee, Diaconis, and Sly '10 [5]):

$$d_{TV}(X_{ij}, \mathrm{Geom}(C)) \to 0 \quad \text{as } n \to \infty$$

Asymptotically independent entries



- Empirical distribution of eigenvalues $\Rightarrow$ circular law (Nguyen '14 [12])

Smooth margins: $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$ so that $\frac{\max \mathbf{a}}{\min \mathbf{a}}, \frac{\max \mathbf{b}}{\min \mathbf{b}} \leq \phi = (1+\sqrt{5})/2 \approx 1.618$.

- Polynomial time approximate algorithm for computing $\mathrm{T}(\mathbf{a}, \mathbf{b})$

Outline

Introduction

Indpendence heuristic and second-order phase transition

Barvinok's conjecture and first-order phase transition

Typical table

Strong duality between typical table and MLE

Conjecture (Independence heuristic, Good '50)

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$$

where

$$\mathrm{G}(\mathbf{a},\mathbf{b}) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{a_i+n-1}{n-1} \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}.$$

Conjecture (Independence heuristic, Good '50)

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$$

where

$$\mathrm{G}(\mathbf{a},\mathbf{b}) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{a_i+n-1}{n-1} \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}.$$

Reasoning:

- $X \sim \text{Uniform}\,(\mathcal{S}_N)$, $\mathcal{S}_N := \{\text{CT's with total sum } N = \sum a_i = \sum b_j\}$

Conjecture (Independence heuristic, Good '50)

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$$

where

$$\mathrm{G}(\mathbf{a},\mathbf{b}) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{a_i+n-1}{n-1} \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}.$$

Reasoning:

- $X \sim \text{Uniform}(\mathcal{S}_N)$, $\mathcal{S}_N := \{\text{CT's with total sum } N = \sum a_i = \sum b_j\}$
- $\mathcal{R}_n(\mathbf{a}) := \{X \text{ has row margins } \mathbf{a}\}$, $\quad \mathcal{C}_m(\mathbf{b}) := \{X \text{ has column margins } \mathbf{b}\}$.

Conjecture (Independence heuristic, Good '50)

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$$

where

$$\mathrm{G}(\mathbf{a},\mathbf{b}) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{a_i+n-1}{n-1} \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}.$$

Reasoning:

- $X \sim \text{Uniform}\,(\mathcal{S}_N)$, $\mathcal{S}_N := \{\text{CT's with total sum } N = \sum a_i = \sum b_j\}$
- $\mathcal{R}_n(\mathbf{a}) := \{X \text{ has row margins } \mathbf{a}\}$, $\quad \mathcal{C}_m(\mathbf{b}) := \{X \text{ has column margins } \mathbf{b}\}$.
- $\mathbb{P}(\mathcal{R}_n(\mathbf{r}) \cap \mathcal{C}_m(\mathbf{c})) = \dfrac{\mathrm{T}(\mathbf{a},\mathbf{b})}{|\mathcal{S}_N|}$, $\quad \mathbb{P}(\mathcal{R}_n(\mathbf{r})) = \dfrac{|\mathcal{R}_n(\mathbf{r})|}{|\mathcal{S}_N|}$, $\quad \mathbb{P}(\mathcal{C}_n(\mathbf{c})) = \dfrac{|\mathcal{C}_n(\mathbf{c})|}{|\mathcal{S}_N|}$

Conjecture (Independence heuristic, Good '50)

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$$

where

$$\mathrm{G}(\mathbf{a},\mathbf{b}) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{a_i+n-1}{n-1} \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}.$$

Reasoning:

- $X \sim \text{Uniform}\,(\mathcal{S}_N)$, $\mathcal{S}_N := \{\text{CT's with total sum } N = \sum a_i = \sum b_j\}$
- $\mathcal{R}_n(\mathbf{a}) := \{X \text{ has row margins } \mathbf{a}\}, \quad \mathcal{C}_m(\mathbf{b}) := \{X \text{ has column margins } \mathbf{b}\}$.
- $\mathbb{P}(\mathcal{R}_n(\mathbf{r}) \cap \mathcal{C}_m(\mathbf{c})) = \dfrac{\mathrm{T}(\mathbf{a},\mathbf{b})}{|\mathcal{S}_N|}, \quad \mathbb{P}(\mathcal{R}_n(\mathbf{r})) = \dfrac{|\mathcal{R}_n(\mathbf{r})|}{|\mathcal{S}_N|}, \quad \mathbb{P}(\mathcal{C}_n(\mathbf{c})) = \dfrac{|\mathcal{C}_n(\mathbf{c})|}{|\mathcal{S}_N|}$
- $|\mathcal{S}_N| = \binom{N+mn-1}{mn-1}$, $|\mathcal{R}_n(\mathbf{a})| = \prod_{i=1}^{m} \binom{a_i+n-1}{n-1}$, $|\mathcal{C}_m(\mathbf{b})| = \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}$

Conjecture (Independence heuristic, Good '50)

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$$

where

$$\mathrm{G}(\mathbf{a},\mathbf{b}) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{a_i+n-1}{n-1} \prod_{j=1}^{n} \binom{b_j+m-1}{m-1}.$$

Reasoning:

- $X \sim \text{Uniform}\,(\mathcal{S}_N)$, $\mathcal{S}_N := \{\text{CT's with total sum } N = \sum a_i = \sum b_j\}$
- $\mathcal{R}_n(\mathbf{a}) := \{X \text{ has row margins } \mathbf{a}\}, \quad \mathcal{C}_m(\mathbf{b}) := \{X \text{ has column margins } \mathbf{b}\}$.
- $\mathbb{P}(\mathcal{R}_n(\mathbf{r}) \cap \mathcal{C}_m(\mathbf{c})) = \dfrac{\mathrm{T}(\mathbf{a},\mathbf{b})}{|\mathcal{S}_N|}, \quad \mathbb{P}(\mathcal{R}_n(\mathbf{r})) = \dfrac{|\mathcal{R}_n(\mathbf{r})|}{|\mathcal{S}_N|}, \quad \mathbb{P}(\mathcal{C}_n(\mathbf{c})) = \dfrac{|\mathcal{C}_n(\mathbf{c})|}{|\mathcal{S}_N|}$
- $|\mathcal{S}_N| = \dbinom{N+mn-1}{mn-1}$, $|\mathcal{R}_n(\mathbf{a})| = \displaystyle\prod_{i=1}^{m} \binom{a_i+n-1}{n-1}$, $|\mathcal{C}_m(\mathbf{b})| = \displaystyle\prod_{j=1}^{n} \binom{b_j+m-1}{m-1}$
- $$\frac{\mathbb{P}(\mathcal{R}_n(\mathbf{a}) \cap \mathcal{C}_m(\mathbf{b}))}{\mathbb{P}(\mathcal{R}_n(\mathbf{a}))\,\mathbb{P}(\mathcal{C}_m(\mathbf{b}))} = \frac{\mathrm{T}(\mathbf{a},\mathbf{b})}{\mathrm{G}(\mathbf{a},\mathbf{b})}$$

History of the Independence Heuristic (IH) $\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$:

History of the Independence Heuristic (IH) $\mathrm{T}(\mathbf{a}, \mathbf{b}) \approx \mathrm{G}(\mathbf{a}, \mathbf{b})$**:**

- Given implicitly by Good in 1963 [10] and later formally in 1963 [8] and 1976 [9]

History of the Independence Heuristic (IH) $\mathrm{T}(\mathbf{a}, \mathbf{b}) \approx \mathrm{G}(\mathbf{a}, \mathbf{b})$:

- Given implicitly by Good in 1963 [10] and later formally in 1963 [8] and 1976 [9]
- Experimentally verified by Good and Crook [7] in 1977 and Diagonis and Gangolli [6] in 1995

History of the Independence Heuristic (IH) $\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$:

- Given implicitly by Good in 1963 [10] and later formally in 1963 [8] and 1976 [9]
- Experimentally verified by Good and Crook [7] in 1977 and Diagonis and Gangolli [6] in 1995
- In 2008, Greenhill and MacKay [11] proved that

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \sim \sqrt{e}\,\mathrm{G}(\mathbf{a},\mathbf{b})$$

for small margins: $\max(a_1,\dots,a_m)\cdot\max(b_1,\dots,b_n) = O(N^{2/3})$

History of the Independence Heuristic (IH) $\mathrm{T}(\mathbf{a},\mathbf{b}) \approx \mathrm{G}(\mathbf{a},\mathbf{b})$:

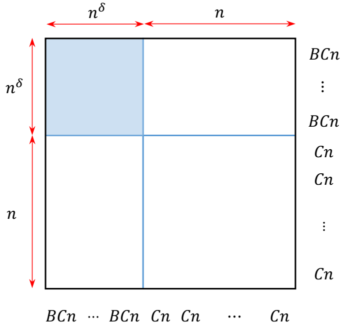
- Given implicitly by Good in 1963 [10] and later formally in 1963 [8] and 1976 [9]
- Experimentally verified by Good and Crook [7] in 1977 and Diagonis and Gangolli [6] in 1995
- In 2008, Greenhill and MacKay [11] proved that

$$\mathrm{T}(\mathbf{a},\mathbf{b}) \sim \sqrt{e}\,\mathrm{G}(\mathbf{a},\mathbf{b})$$

  for small margins: $\max(a_1,\dots,a_m)\cdot\max(b_1,\dots,b_n) = O(N^{2/3})$

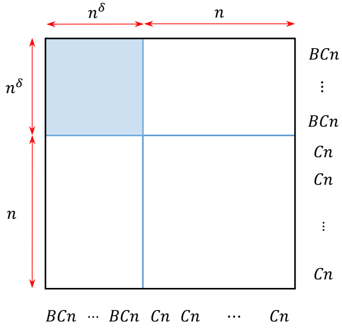
- In 2010, Greenhill and MacKay [4] proved (1) for uniform linear margins $n=m$, $\mathbf{a}=\mathbf{b}=(Cn, Cn, \dots, Cn)$, $C>0$

- Two margins: $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{(n - n^\delta)})$, $0 \le \delta \le 1$



Two margins

- Two margins: $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{(n-n^\delta)})$, $0 \le \delta \le 1$



- IH undercounts: For $\delta = 1$, Barvinok [1] shows that

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathrm{T}(\mathbf{a}, \mathbf{b}) > \lim_{n\to\infty} \frac{1}{n^2} \log \mathrm{G}(\mathbf{a}, \mathbf{b}).$$

In other words, the rows and columns of CTs attract each other

Theorem (L., and Pak '22)

Let $0 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots\ldots, Cn}^{n - n^\delta}) \in \mathbb{N}^n$*.*
Let $B_c := 1 + \sqrt{1 + 1/C}$ *and* $f(x) := (x+1)\log(x+1) - x\log x$*.*

Theorem (L., and Pak '22)

Let $0<\delta<1$ *and* $\mathbf{a}=\mathbf{b}=(\overbrace{BCn,\ldots,BCn}^{n^\delta},\overbrace{Cn,\ldots\ldots\ldots,Cn}^{n-n^\delta})\in\mathbb{N}^n$.
Let $B_c:=1+\sqrt{1+1/C}$ *and* $f(x):=(x+1)\log(x+1)-x\log x$.

(i) $\displaystyle\lim_{n\to\infty}\frac{1}{n^2}\log\mathrm{T}(\mathbf{a},\mathbf{b})=\lim_{n\to\infty}\frac{1}{n^2}\log\mathrm{G}(\mathbf{a},\mathbf{b})=f(C)$

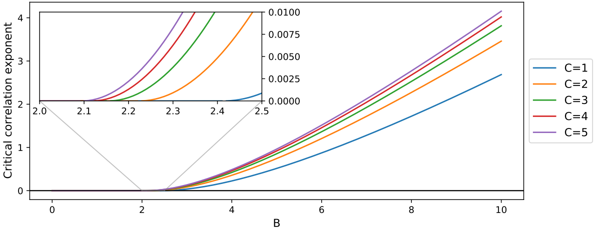
Theorem (L., and Pak '22)

Let $0<\delta<1$ *and* $\mathbf{a}=\mathbf{b}=(\overbrace{BCn,\dots,BCn}^{n^\delta},\overbrace{Cn,\dots\dots,Cn}^{n-n^\delta})\in\mathbb{N}^n$.
Let $B_c := 1+\sqrt{1+1/C}$ *and* $f(x):=(x+1)\log(x+1)-x\log x$.

(i) $\displaystyle\lim_{n\to\infty}\frac{1}{n^2}\log\mathrm{T}(\mathbf{a},\mathbf{b})=\lim_{n\to\infty}\frac{1}{n^2}\log\mathrm{G}(\mathbf{a},\mathbf{b})=f(C)$

(ii)

$$\lim_{n\to\infty}\frac{1}{n^{1+\delta}}\log\frac{\mathrm{T}(\mathbf{a},\mathbf{b})}{\mathrm{G}(\mathbf{a},\mathbf{b})}=\begin{cases}0 & \text{if } B\le B_c\\ C(B-B_c)\log\left(1+\frac{1}{C}\right)-2\big(f(BC)-f(B_cC)\big)>0 & \text{if } B>B_c.\end{cases}$$



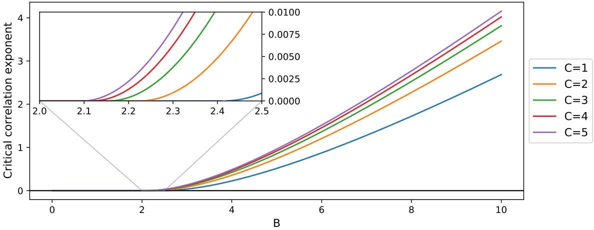
Theorem (L., and Pak '22)

Let $0 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$.
Let $B_c := 1 + \sqrt{1 + 1/C}$ *and* $f(x) := (x+1)\log(x+1) - x\log x$.

(i) $\displaystyle\lim_{n\to\infty} \frac{1}{n^2} \log \mathrm{T}(\mathbf{a}, \mathbf{b}) = \lim_{n\to\infty} \frac{1}{n^2} \log \mathrm{G}(\mathbf{a}, \mathbf{b}) = f(C)$

(ii)

$$\lim_{n\to\infty} \frac{1}{n^{1+\delta}} \log \frac{\mathrm{T}(\mathbf{a}, \mathbf{b})}{\mathrm{G}(\mathbf{a}, \mathbf{b})} = \begin{cases} 0 & \text{if} \quad B \le B_c \\ C(B - B_c)\log\left(1 + \frac{1}{C}\right) - 2\big(f(BC) - f(B_cC)\big) > 0 & \text{if} \quad B > B_c. \end{cases}$$



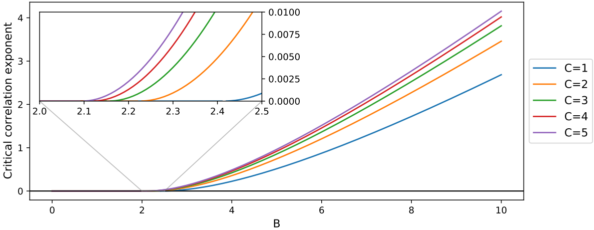
- Asmptotic independence $\xrightarrow{B\nearrow}$ Positive correlation

Theorem (L., and Pak '22)

Let $0 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$.
Let $B_c := 1 + \sqrt{1 + 1/C}$ *and* $f(x) := (x+1)\log(x+1) - x \log x$.

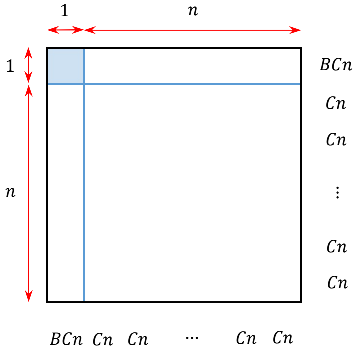
(ii)

$$\lim_{n\to\infty} \frac{1}{n^{1+\delta}} \log \frac{\mathrm{T}(\mathbf{a},\mathbf{b})}{\mathrm{G}(\mathbf{a},\mathbf{b})} = \begin{cases} 0 & \text{if } \; B \le B_c \\ C(B - B_c)\log\left(1 + \frac{1}{C}\right) - 2\big(f(BC) - f(B_cC)\big) > 0 & \text{if } \; B > B_c. \end{cases}$$

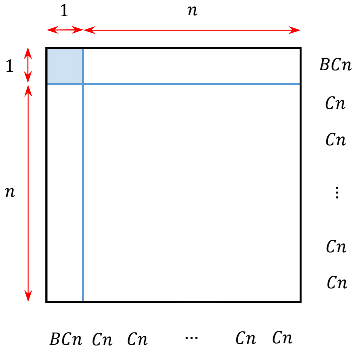


- Asmptotic independence $\xrightarrow{B\nearrow}$ Positive correlation
- Where is this phase transition coming from?

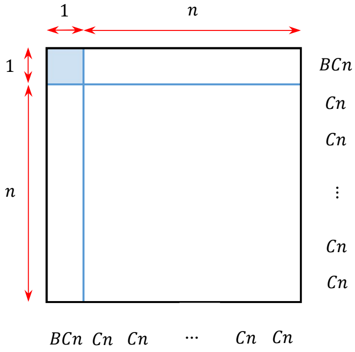
Outline



- Let $\mathbf{a} = \mathbf{b} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^{n+1}$. Let $X = (X_{ij})$ be the uniform contingency table with this margin.

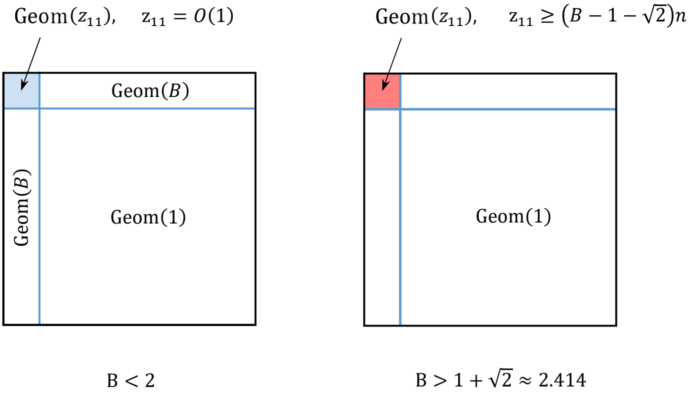


- Let $\mathbf{a} = \mathbf{b} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^{n+1}$. Let $X = (X_{ij})$ be the uniform contingency table with this margin.
- Do we still have convergence to geometric entries for all $B, C \geq 1$?

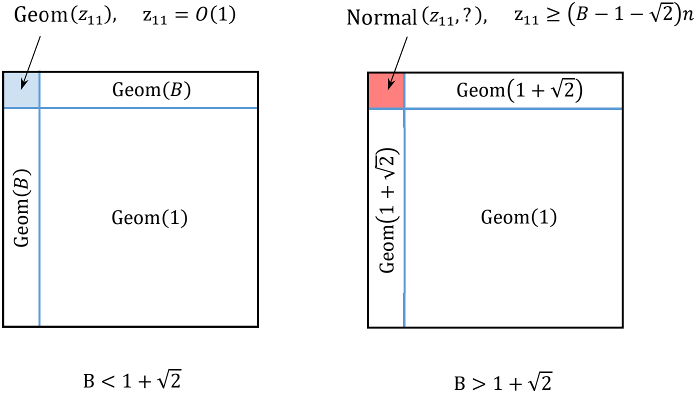


- Let $\mathbf{a} = \mathbf{b} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^{n+1}$. Let $X = (X_{ij})$ be the uniform contingency table with this margin.
- Do we still have convergence to geometric entries for all $B, C \geq 1$?
- If so, what are the means of the geometric distribution in each block?

- Based on his **typical table** computation, Barvinok conjectured in 2010 that each entry in X is asymptotically distributed as a geometric variable;
- Furthermore, for $C = 1$, he conjecture that $\mathbb{E}[X_{11}] = O(1)$ for $B < 2$ and $\mathbb{E}[X_{11}] = \Theta(n)$ for $B > 1 + \sqrt{2}$.



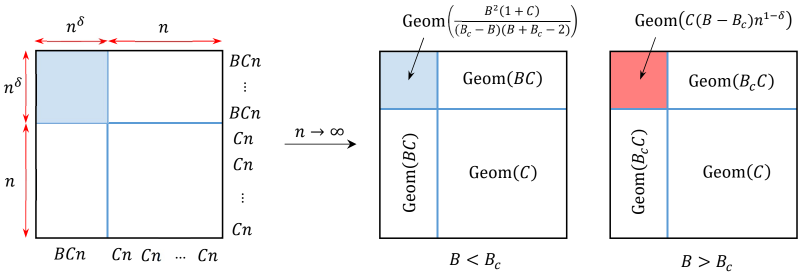
- In 2018, Dittmer and Pak tested Barvinok's conjecture using a new MCMC algorithm (Jerrum's Burnside process) to sample a uniform contingency table of reasonable size
- They conjectured that $B_c = 1 + \sqrt{2}$ is the critical value and X_{11} actually converges to a normal variable with growing mean



Simulation of corner entry X_{11}

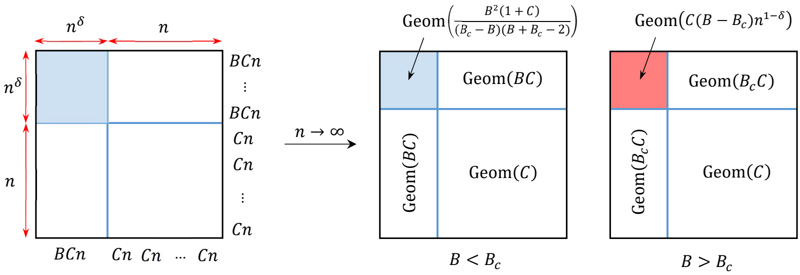
Theorem (Dittmer, L., and Pak '20)

Let $1/2 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$*. Let* $B_c := 1 + \sqrt{1 + 1/C}$ *and* $X \sim \mathrm{Uniform}(\mathcal{T}(\mathbf{a}, \mathbf{b}))$*. Then X marginally converges to the following matrix in total variation distance:*



Theorem (Dittmer, L., and Pak '20)

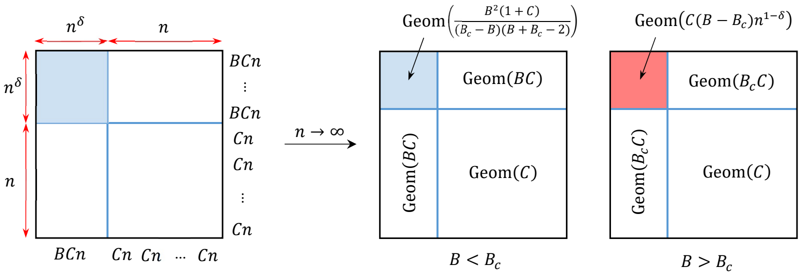
Let $1/2 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$*. Let* $B_c := 1 + \sqrt{1 + 1/C}$ *and* $X \sim \mathrm{Uniform}(\mathcal{T}(\mathbf{a}, \mathbf{b}))$*. Then* X *marginally converges to the following matrix in total variation distance:*



- We also show polynomial rate of convergence in d_{TV}.

Theorem (Dittmer, L., and Pak '20)

Let $1/2 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$. *Let* $B_c := 1 + \sqrt{1 + 1/C}$ *and* $X \sim \mathrm{Uniform}(\mathcal{T}(\mathbf{a}, \mathbf{b}))$. *Then* X *marginally converges to the following matrix in total variation distance:*



- We also show polynomial rate of convergence in d_{TV}.
- But where is this phase transition coming from?

Outline

Definition

Fix margins $\mathbf{a}, \mathbf{c} \in \mathbb{N}^n$. *Let* $\mathcal{P}(\mathbf{a}, \mathbf{b}) \subseteq \mathbb{R}^{n\times n}_{\geq 0}$ *denote the set of all matrices with non-negative real entries with margins* $\mathbf{r}$ *and* $\mathbf{c}$. *For each* $X = (x_{ij}) \in \mathcal{P}(\mathbf{a}, \mathbf{b})$, *define*

$$g(X) = \sum_{1\leq i,j\leq n} (x_{ij}+1)\log(x_{ij}+1) - x_{ij}\log(x_{ij}).$$

The typical table $Z \in \mathcal{P}(\mathbf{a}, \mathbf{b})$ *for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$ *is defined by*

$$Z = \arg\max_{X\in\mathcal{P}(\mathbf{a},\mathbf{b})} g(X).$$

Theorem (Barvinok '09, '10)

Fix any margins $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$*. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*. Let* $N = \sum_{i=1}^{m} a_i = \sum_{j=1}^{m} b_j$ *denote the total sum.*

Theorem (Barvinok '09, '10)

Fix any margins $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$*. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*. Let* $N = \sum_{i=1}^{m} a_i = \sum_{j=1}^{m} b_j$ *denote the total sum.*

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

$$g(Z) - \gamma(m+n)\log N \leq \log \mathrm{T}(\mathbf{a}, \mathbf{b}) \leq g(Z),$$

Theorem (Barvinok '09, '10)

Fix any margins $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$*. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*. Let* $N = \sum_{i=1}^m a_i = \sum_{j=1}^m b_j$ *denote the total sum.*

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

$$g(Z) - \gamma(m+n)\log N \leq \log \mathrm{T}(\mathbf{a}, \mathbf{b}) \leq g(Z),$$

(ii) *Let* $Y = (Y_{ij})$ *be the* $(n \times n)$ *random matrix of independent entries,* $Y_{ij} \sim \mathrm{Geom}(z_{ij})$*. Then* Y *is uniform on* $\mathcal{T}(\mathbf{a}, \mathbf{b})$ *conditional on being in* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*.*

Theorem (Barvinok '09, '10)

Fix any margins $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$*. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*. Let* $N = \sum_{i=1}^m a_i = \sum_{j=1}^m b_j$ *denote the total sum.*

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

$$g(Z) - \gamma(m+n)\log N \leq \log \mathrm{T}(\mathbf{a}, \mathbf{b}) \leq g(Z),$$

(ii) *Let* $Y = (Y_{ij})$ *be the* $(n \times n)$ *random matrix of independent entries,* $Y_{ij} \sim \mathrm{Geom}(z_{ij})$*. Then* Y *is uniform on* $\mathcal{T}(\mathbf{a}, \mathbf{b})$ *conditional on being in* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*.*

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

$$\mathbb{P}(Y \in \mathcal{T}(\mathbf{a}, \mathbf{b})) = e^{-g(Z)}\mathrm{T}(\mathbf{a}, \mathbf{b}) \geq N^{-\gamma n}.$$

Theorem (Barvinok '09, '10)

Fix any margins $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$*. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*. Let* $N = \sum_{i=1}^m a_i = \sum_{j=1}^m b_j$ *denote the total sum.*

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

$$g(Z) - \gamma(m+n)\log N \leq \log \mathrm{T}(\mathbf{a}, \mathbf{b}) \leq g(Z),$$

Theorem (Barvinok '09, '10)

Fix any margins $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$*. Let* $Z = (z_{ij})$ *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$*. Let* $N = \sum_{i=1}^m a_i = \sum_{j=1}^m b_j$ *denote the total sum.*

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

$$g(Z) - \gamma(m+n) \log N \le \log \mathrm{T}(\mathbf{a}, \mathbf{b}) \le g(Z),$$

- For **(i)**: Lower bound is hard; Upper bound is immediate from the GF:

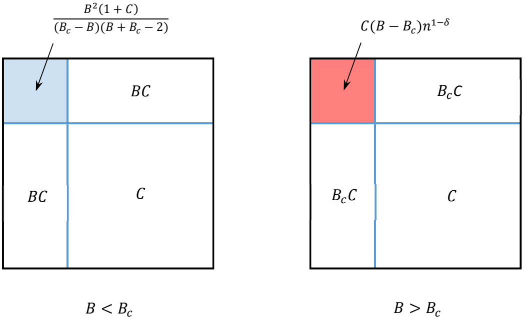
$$\prod_{i=1}^n \prod_{j=1}^n \frac{1}{1 - x_i y_j} = \sum_{\mathbf{a} \in \mathbb{N}^m, \, \mathbf{b} \in \mathbb{N}^n} \mathrm{T}(\mathbf{a}, \mathbf{b}) \prod_{i=1}^m x_i^{a_i} \prod_{j=1}^m y_j^{b_j}$$

$$\mathrm{T}(\mathbf{a}, \mathbf{b}) \le \inf \left[\prod_{i=1}^m x_i^{a_i} \prod_{j=1}^m y_j^{b_j} \right]^{-1} \prod_{j=1}^n \frac{1}{1 - x_i y_j}$$

$$= \exp\left(\sup \left[\sum_i a_i \log x_i + \sum_j b_j \log y_j + \sum_{ij} \log(1 - x_i y_j) \right] \right) = \exp(g(Z))$$

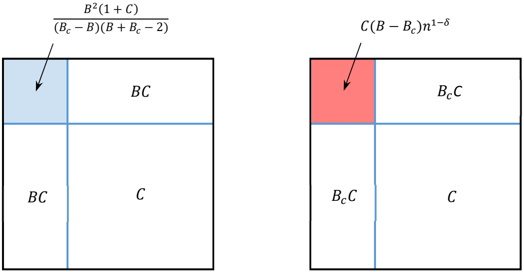
Lemma (Dittmer, L., and Pak '19+)

Let $0 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$.
Let Z *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$. *Let* $B_c := 1 + \sqrt{1 + 1/C}$. *Then for* $0 \le \delta < 1$, *the first order asymptotics of the entries of* Z *are given by:*



Lemma (Dittmer, L., and Pak '19+)

Let $0 < \delta < 1$ *and* $\mathbf{a} = \mathbf{b} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{n-n^\delta}) \in \mathbb{N}^n$.
Let Z *be the typical table for* $\mathcal{T}(\mathbf{a}, \mathbf{b})$. *Let* $B_c := 1 + \sqrt{1 + 1/C}$. *Then for* $0 \leq \delta < 1$, *the first order asymptotics of the entries of* Z *are given by:*



$B < B_c$ $\qquad$ $B > B_c$

- We also show polynomial rate of convergence ← Crucial in volumn phase transition

The (α, β) model

- Can we contruct a parameterized statistical model for $m \times n$ contingency tables?

The (α, β) model

- Can we contruct a parameterized statistical model for $m \times n$ contingency tables?
- Paramters: $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathbf{r}^m$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m) \in \mathbf{r}^n$,

The (α, β) model

- Can we contruct a parameterized statistical model for $m \times n$ contingency tables?
- Paramters: $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m) \in \mathbf{r}^m$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m) \in \mathbf{r}^n$,
- $Y = (Y_{ij})_{i,j}$ a random $m \times n$ matrix in $[0, B]^{m \times n}$ s.t.

$$\mathbb{E}[Y_{ij}] = \psi'(\alpha_i + \beta_j)$$

The (α, β) model

- Can we contruct a parameterized statistical model for $m \times n$ contingency tables?
- Paramters: $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m) \in \mathbf{r}^m$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m) \in \mathbf{r}^n$,
- $Y = (Y_{ij})_{i,j}$ a random $m \times n$ matrix in $[0, B]^{m \times n}$ s.t.

$$\mathbb{E}[Y_{ij}] = \psi'(\alpha_i + \beta_j)$$

- The 'mean function' ψ' comes from an exponential family:
  - μ a base probability measure on interval $[0, B]$
  - For $\theta \in \mathbf{r}$, let μ_θ denote the titled probability measure:

$$\frac{d\mu_\theta}{d\mu}(x) = e^{\theta x - \psi(\theta)}, \quad \psi(\theta) := \log \int_A^B e^{\theta x} d\mu(x).$$

  - Then $\mathbb{E}_{\mu_\theta}[X] = \psi'(\theta)$ and $\mathrm{Var}_{\mu_\theta}(X) = \psi''(\theta) > 0$.

MLE for the (α, β) model

- Suppose we observe a CT $X = (x_{ij})$ with margin $(\mathbf{r}, \mathbf{c})$.

- Suppose we observe a CT $X = (x_{ij})$ with margin $(\mathbf{r}, \mathbf{c})$.
- What is the likelihood of observing X under the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model?

$$G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{i=1}^{n} r_i \alpha_i + \sum_{j=1}^{m} c_j \beta_j - \sum_{i,j} \psi(\alpha_i + \beta_j).$$

(depends only on the margin)

MLE for the (α, β) model

- Suppose we observe a CT $X = (x_{ij})$ with margin $(\mathbf{r}, \mathbf{c})$.
- What is the likelihood of observing X under the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model?

$$G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{i=1}^{n} r_i \alpha_i + \sum_{j=1}^{m} c_j \beta_j - \sum_{i,j} \psi(\alpha_i + \beta_j).$$

(depends only on the margin)

- The MLE for the margin $(\mathbf{r}, \mathbf{c})$:

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \arg\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- Suppose we observe a CT $X = (x_{ij})$ with margin $(\mathbf{r}, \mathbf{c})$.
- What is the likelihood of observing X under the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model?

$$G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{i=1}^{n} r_i \alpha_i + \sum_{j=1}^{m} c_j \beta_j - \sum_{i,j} \psi(\alpha_i + \beta_j).$$

(depends only on the margin)

- The MLE for the margin $(\mathbf{r}, \mathbf{c})$:

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \operatorname*{arg\,max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- MLE is not unique:

$$(\boldsymbol{\alpha}, \boldsymbol{\beta})\text{-model} \overset{d}{=} (\boldsymbol{\alpha}', \boldsymbol{\beta}')\text{-model} \quad \Longleftrightarrow \quad \exists \delta \in \mathbf{r} \text{ s.t. } \boldsymbol{\alpha}' - \boldsymbol{\alpha} \equiv \delta \equiv \boldsymbol{\beta} - \boldsymbol{\beta}'$$

MLE for the (α, β) model

- Suppose we observe a CT $X = (x_{ij})$ with margin $(\mathbf{r}, \mathbf{c})$.
- What is the likelihood of observing X under the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model?

$$G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{i=1}^{n} r_i \alpha_i + \sum_{j=1}^{m} c_j \beta_j - \sum_{i,j} \psi(\alpha_i + \beta_j).$$

(depends only on the margin)

- The MLE for the margin $(\mathbf{r}, \mathbf{c})$:

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \arg\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} G^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- MLE is not unique:

$$(\boldsymbol{\alpha}, \boldsymbol{\beta})\text{-model} \overset{d}{=} (\boldsymbol{\alpha}', \boldsymbol{\beta}')\text{-model} \iff \exists \delta \in \mathbf{r} \text{ s.t. } \boldsymbol{\alpha}' - \boldsymbol{\alpha} \equiv \delta \equiv \boldsymbol{\beta} - \boldsymbol{\beta}'$$

- $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is an MLE for margin $(\mathbf{r}, \mathbf{c})$ iff the expected table $\mathbb{E}[\mu_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}]$ satisfies the margin $(\mathbf{r}, \mathbf{c})$

Definition (Generalized typical table)

Fix margins $\mathbf{r} = (r_1, \dots, r_m) \in \mathbf{r}^m$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbf{r}^n$. For each $X = (x_{ij}) \in \mathcal{T}(\mathbf{r}, \mathbf{c})$,

$$g(X) := \sum_{1 \le i \le m,\, 1 \le j \le n} f(x_{ij}),$$

where the function $f\colon [0, B] \to [-\infty, 0]$ is defined by

$$f(x) = -D(\phi(x) \,\|\, 0) = \begin{cases} \log\, \mu(\{0\}) & \text{if } x = 0 \\ \psi(\phi(x)) - x\phi(x) & \text{for } x \in (0, B) \\ \log\, \mu(\{B\}) & \text{if } x = B. \end{cases}$$

The *(generalized) typical table* Z for margin $(\mathbf{r}, \mathbf{c})$ is defined by

$$Z = \operatorname{arg\,max}_{X \in \mathcal{P}(\mathbf{r},\mathbf{c})} g(X)$$

- $Z =$ matrix with margin $(\mathbf{r}, \mathbf{c})$ that has minimum total KL-divergence from the base measure $\mu_{0,0}$.
- Unique sol. of a strongly concave maximization problem in a compact domain

Lemma (Strong duality between MLE and typical table; L. Mukherjee '23+)

Fix margins $\mathbf{r} = (r_1, \ldots, r_m) \in [0, nB]^m$, $\mathbf{c} = (c_1, \ldots, c_n) \in [0, mB]^n$. *Then*

(i) *If* $Z = (z_{ij})$ *is the typical table for* $\mathcal{T}(\mathbf{r}, \mathbf{c})$ *and satisfies* $0 < Z_{ij} < B$ *for all* i, j, *then there exists an MLE* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *for margin* $(\mathbf{r}, \mathbf{c})$ *which satisfies*

$$z_{ij} = \psi'(\alpha_i + \beta_j) \quad \text{for all } 1 \leq i \leq m \text{ and } 1 \leq j \leq n,$$

and $\|\boldsymbol{\alpha}\|_\infty, \|\boldsymbol{\beta}\|_\infty \leq C$, *where* C *is a finite constant depending only on* (δ, μ). *In particular, an MLE* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *for margin* $(\mathbf{r}, \mathbf{c})$ *exists.*

Lemma (Strong duality between MLE and typical table; L. Mukherjee '23+)

Fix margins $\mathbf{r} = (r_1, \dots, r_m) \in [0, nB]^m$, $\mathbf{c} = (c_1, \dots, c_n) \in [0, mB]^n$. *Then*

(i) *If* $Z = (z_{ij})$ *is the typical table for* $\mathcal{T}(\mathbf{r}, \mathbf{c})$ *and satisfies* $0 < Z_{ij} < B$ *for all* i, j, *then there exists an MLE* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *for margin* $(\mathbf{r}, \mathbf{c})$ *which satisfies*

$$z_{ij} = \psi'(\alpha_i + \beta_j) \quad \text{for all } 1 \le i \le m \text{ and } 1 \le j \le n,$$

and $\|\boldsymbol{\alpha}\|_\infty, \|\boldsymbol{\beta}\|_\infty \le C$, *where* C *is a finite constant depending only on* (δ, μ). *In particular, an MLE* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *for margin* $(\mathbf{r}, \mathbf{c})$ *exists.*

(ii) *For each* $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{R}^m \times \mathbb{R}^n$, *define* $X^{\boldsymbol{\alpha}, \boldsymbol{\beta}} = (x_{ij})_{i,j} \in \mathbb{R}^{m \times n}$ *by*

$$x_{ij} := \psi'(\alpha_i + \beta_j) \quad \text{for all } 1 \le i \le m \text{ and } 1 \le j \le n.$$

Suppose $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *is an MLE for margin* $(\mathbf{r}, \mathbf{c})$. *Then*

$$\sup_{X \in \mathcal{T}(\mathbf{r}, \mathbf{c})} g(X) = g(X^{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}) = -G^{\mathbf{r}, \mathbf{c}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = -\sup_{\boldsymbol{\alpha}, \boldsymbol{\beta}} G^{\mathbf{r}, \mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

In particular, $X^{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}$ *is a typical table for margin* $(\mathbf{r}, \mathbf{c})$.

Thanks a lot!

[1] Alexander Barvinok. "Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes". In: *International Mathematics Research Notices* 2009.2 (2009), pp. 348–385.

[2] Alexander Barvinok. *Combinatorics and complexity of partition functions*. Vol. 9. Springer, 2016.

[3] Petter Brändén, Jonathan Leake, and Igor Pak. "Lower bounds for contingency tables via Lorentzian polynomials". In: *arXiv preprint arXiv:2008.05907* (2020).

[4] E Rodney Canfield and Brendan D McKay. "Asymptotic enumeration of integer matrices with large equal row and column sums". In: *Combinatorica* 30.6 (2010), p. 655.

[5] Sourav Chatterjee, Persi Diaconis, and Allan Sly. "Properties of uniform doubly stochastic matrices". In: *arXiv preprint arXiv:1010.6136* (2010).

[6] Persi Diaconis and Anil Gangolli. "Rectangular arrays with fixed margins". In: *Discrete probability and algorithms*. Springer, 1995, pp. 15–41.

[7] IJ Good and JF Crook. "The enumeration of arrays and a generalization related to contingency tables". In: *Discrete Mathematics* 19.1 (1977), pp. 23–45.

[8] Irving J Good. "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables". In: *The Annals of Mathematical Statistics* 34.3 (1963), pp. 911–934.

[9] Irving J Good. "On the application of symmetric Dirichlet distributions and their mixtures to contingency tables". In: *The Annals of Statistics* 4.6 (1976), pp. 1159–1189.

[10] Isidore Jacob Good. *Probability and the Weighing of Evidence*. Tech. rep. C. Griffin London, 1950.

[11] Catherine Greenhill and Brendan D McKay. "Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums". In: *Advances in Applied Mathematics* 41.4 (2008), pp. 459–481.

[12] Hoi H Nguyen. "Random doubly stochastic matrices: the circular law". In: *Annals of Probability* 42.3 (2014), pp. 1161–1196.

[13] Austin Shapiro. "Bounds on the number of integer points in a polytope via concentration estimates". In: *arXiv preprint arXiv:1011.6252* (2010).