Large random matrices with given margins

Hanbaek Lyu

University of Wisconsin - Madison

Based on joint work with Sumit Mukherjee (Columbia)

Jan. 16, 2016

- *(Base model)* $\mu =$ probability measure on $\mathbb{Z}_{\geq 0}$, and let

$$A := \inf\{\mathrm{supp}(\mu)\} \leq \sup\{\mathrm{supp}(\mu)\} =: B.$$

$X \sim \mu^{\otimes(m\times n)}$: $(m \times n)$ random matrix with i.i.d. entries drawn from μ

- (*Base model*) $\mu =$ probability measure on $\mathbb{Z}_{\geq 0}$, and let

$$A := \inf\{\mathrm{supp}(\mu)\} \leq \sup\{\mathrm{supp}(\mu)\} =: B.$$

$X \sim \mu^{\otimes(m \times n)}$: $(m \times n)$ random matrix with i.i.d. entries drawn from μ

- (*Margins*) For a matrix $\mathbf{x} = (x_{ij}) \in \mathbb{R}^{m \times n}$, $(r(\mathbf{x}), c(\mathbf{x}))$ = margin of $\mathbf{x}$:

$$r(\mathbf{x}) := (r_1(\mathbf{x}), \dots, r_m(\mathbf{x})); \quad r_i(\mathbf{x}) := x_{i1} + \cdots + x_{in} \qquad (\triangleright \textit{ row margin of } \mathbf{x})$$
$$c(\mathbf{x}) := (c_1(\mathbf{x}), \dots, c_n(\mathbf{x})); \quad c_j(\mathbf{x}) := x_{1j} + \cdots + x_{mj} \qquad (\triangleright \textit{ column margin of } \mathbf{x})$$

$$\mathcal{T}(\mathbf{r}, \mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m \times n} \,:\, r(\mathbf{x}) = \mathbf{r},\; c(\mathbf{x}) = \mathbf{c} \right\}$$

- *(Base model)* μ = probability measure on $\mathbb{Z}_{\geq 0}$, and let

$$A := \inf\{\mathrm{supp}(\mu)\} \leq \sup\{\mathrm{supp}(\mu)\} =: B.$$

$X \sim \mu^{\otimes(m\times n)}$: $(m \times n)$ random matrix with i.i.d. entries drawn from μ

- *(Margins)* For a matrix $\mathbf{x} = (x_{ij}) \in \mathbb{R}^{m\times n}$, $(r(\mathbf{x}), c(\mathbf{x}))$ = margin of $\mathbf{x}$:

$$r(\mathbf{x}) := (r_1(\mathbf{x}), \dots, r_m(\mathbf{x})); \quad r_i(\mathbf{x}) := x_{i1} + \cdots + x_{in} \qquad (\triangleright \textit{ row margin of } \mathbf{x})$$
$$c(\mathbf{x}) := (c_1(\mathbf{x}), \dots, c_n(\mathbf{x})); \quad c_j(\mathbf{x}) := x_{1j} + \cdots + x_{mj} \qquad (\triangleright \textit{ column margin of } \mathbf{x})$$

$$\mathcal{T}(\mathbf{r}, \mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m\times n} \,:\, r(\mathbf{x}) = \mathbf{r},\ c(\mathbf{x}) = \mathbf{c} \right\}$$

- *(Main question)*

If we condition $X \sim \mu^{\otimes(m\times n)}$ on being in $\mathcal{T}(\mathbf{r}, \mathbf{c})$, how does it look like?

- This question still makes sense if μ is not a probability measure (i.e., counting measure on $\mathbb{Z}_{\geq 0}$)

- (*Main question*)

  If we condition $X \sim \mu^{\otimes(m \times n)}$ on being in $\mathcal{T}(\mathbf{r}, \mathbf{c})$, how does it look like?

- (*Main question*)

  If we condition $X \sim \mu^{\otimes(m\times n)}$ on being in $\mathcal{T}(\mathbf{r},\mathbf{c})$, how does it look like?

- High-level answer:

- (*Main question*)

  If we condition $X \sim \mu^{\otimes(m\times n)}$ on being in $\mathcal{T}(\mathbf{r}, \mathbf{c})$, how does it look like?

- High-level answer:
  - (*Minimum Relative Entropy Perspective*): The expectation of the minimum relative entropy random matrix from the base model constrained to have expected margin $(\mathbf{r}, \mathbf{c})$

- (*Main question*)

  If we condition $X \sim \mu^{\otimes(m \times n)}$ on being in $\mathcal{T}(\mathbf{r}, \mathbf{c})$, how does it look like?

- High-level answer:
  - (*Minimum Relative Entropy Perspective*): The expectation of the minimum relative entropy random matrix from the base model constrained to have expected margin $(\mathbf{r}, \mathbf{c})$

  - (*Maximum Liklihood Perspective*): The expectation of the maximum likelihood entrywise exponential tilting of the base model for margin $(\mathbf{r}, \mathbf{c})$

- μ_θ := exponentially tilted probability measure given by

$$\frac{d\mu_\theta}{d\mu}(x) = e^{\theta x - \psi(\theta)}, \quad \psi(\theta) := \log \int_{\mathbb{R}} e^{\theta x} d\mu(x) = \text{log partition function}$$

  - Set of all allowed tilting parameters:

$$\Theta^\circ := \text{Interior}\left(\left\{\theta \in \mathbb{R} : \int_{\mathbb{R}} e^{\theta x} d\mu(x) < \infty\right\}\right) = \text{Nonempty interval}$$

Exponential tilting

- $\mu_\theta :=$ exponentially tilted probability measure given by

$$\frac{d\mu_\theta}{d\mu}(x) = e^{\theta x - \psi(\theta)}, \quad \psi(\theta) := \log \int_{\mathbb{R}} e^{\theta x} d\mu(x) = \text{log partition function}$$

  - Set of all allowed tilting parameters:

$$\Theta^\circ := \text{Interior}\left(\left\{ \theta \in \mathbb{R} : \int_{\mathbb{R}} e^{\theta x} d\mu(x) < \infty \right\} \right) = \text{Nonempty interval}$$

- Elementary facts: For $\theta \in \Theta^\circ$,

$$\mathbb{E}_{X\sim\mu_\theta}[X] = \psi'(\theta), \quad \text{Var}_{X\sim\mu_\theta}(X) = \psi''(\theta) > 0.$$

  - $\psi' : \Theta^\circ \to (A, B)$ is strictly increasing ($\triangleright$ `tilt2mean` function)
  - $\phi = (\psi')^{-1} : (A, B) \to \Theta^\circ$ is strictly increasing ($\triangleright$ `mean2tilt` function)

Typical table

- For $\theta \in \Theta^\circ$, the **relative entropy** from the base mesure μ to the tilted probability measure μ_θ is

$$D(\mu_\theta \| \mu) := \int_{x\in\mathbb{R}} \log\left(\frac{d\mu_\theta}{d\mu}(x)\right) \, d\mu_\theta(x) = \theta\psi'(\theta) - \psi(\theta).$$

- For $\theta \in \Theta^\circ$, the **relative entropy** from the base mesure μ to the tilted probability measure μ_θ is
$$D(\mu_\theta \| \mu) := \int_{x\in\mathbb{R}} \log\left(\frac{d\mu_\theta}{d\mu}(x)\right) d\mu_\theta(x) = \theta\psi'(\theta) - \psi(\theta).$$
- Fix a $m \times n$ margin $(\mathbf{r}, \mathbf{c}) \in \mathbb{R}^m \times \mathbb{R}^n$. The **typical table** Z for margin $(\mathbf{r}, \mathbf{c})$ is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,min}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m\times n}} \sum_{i,j} \underbrace{D(\mu_{\phi(x_{ij})} \,\|\, \mu)}_{f(x) := D(\mu_{\phi(x)} \,\|\, \mu) = x\,\phi(x) - \psi(\phi(x))}$$

  - Strictly convex objective since $f'(x) = \phi(x)$, $f''(x) = \phi'(x) = \frac{1}{\mathrm{Var}(\mu_{\phi(x)})} > 0$
  - So the typical table $Z^{\mathbf{r},\mathbf{c}}$ is unique if it exists

- For $\theta \in \Theta^\circ$, the **relative entropy** from the base mesure μ to the tilted probability measure μ_θ is

$$D(\mu_\theta \| \mu) := \int_{x\in\mathbb{R}} \log\left(\frac{d\mu_\theta}{d\mu}(x)\right) d\mu_\theta(x) = \theta\psi'(\theta) - \psi(\theta).$$

- Fix a $m \times n$ margin $(\mathbf{r}, \mathbf{c}) \in \mathbb{R}^m \times \mathbb{R}^n$. The **typical table** Z for margin $(\mathbf{r}, \mathbf{c})$ is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,min}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m\times n}} \sum_{i,j} \underbrace{D(\mu_{\phi(x_{ij})} \,\|\, \mu)}_{f(x) := D(\mu_{\phi(x)} \,\|\, \mu) = x\,\phi(x) - \psi(\phi(x))}$$

  - Strictly convex objective since $f'(x) = \phi(x)$, $f''(x) = \phi'(x) = \frac{1}{\mathrm{Var}(\mu_{\phi(x)})} > 0$
  - So the typical table $Z^{\mathbf{r},\mathbf{c}}$ is unique if it exists

- By multivariate Lagrange multipliers, there are 'dual variables' $\boldsymbol{\alpha} \in \mathbb{R}^m$, $\boldsymbol{\beta} \in \mathbb{R}^n$ s.t.

$$Z^{\mathbf{r},\mathbf{c}}_{ij} = \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \quad \text{for all } i, j.$$

  - Dual variable $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ determined by the margin condition:

$$\sum_{i=1}^{m} \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) = \mathbf{r}(i), \qquad \sum_{j=1}^{n} \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) = \mathbf{c}(j) \qquad \forall\, i, j$$

Examples

- $\mu =$ Gaussian

$$\Theta = \mathbb{R}, \qquad (A,B) = (-\infty, \infty), \qquad \psi(\theta) = \frac{\theta^2}{2}, \qquad \psi'(\theta) = \theta, \qquad \phi(x) = x$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = \frac{x^2}{2}$$

$$Z_{ij}^{\mathbf{r},\mathbf{c}} = \frac{\mathbf{r}(i)}{n} + \frac{\mathbf{c}(j)}{m} - \frac{N}{mn} \qquad \left(N = \sum_i \mathbf{r}(i) = \sum_j \mathbf{c}(j)\right)$$

Examples

- $\mu = $ Gaussian

$$\Theta = \mathbb{R}, \qquad (A,B) = (-\infty, \infty), \qquad \psi(\theta) = \frac{\theta^2}{2}, \qquad \psi'(\theta) = \theta, \qquad \phi(x) = x$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = \frac{x^2}{2}$$

$$Z_{ij}^{\mathbf{r},\mathbf{c}} = \frac{\mathbf{r}(i)}{n} + \frac{\mathbf{c}(j)}{m} - \frac{N}{mn} \qquad \left(N = \sum_i \mathbf{r}(i) = \sum_j \mathbf{c}(j)\right)$$

- $\mu = $ Poisson

$$\Theta = \mathbb{R}, \quad (A,B) = (0, \infty), \quad \psi(\theta) = e^{\theta}, \quad \psi'(\theta) = e^{\theta}, \quad \phi(x) = \log x$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = x \log x - x$$

$$Z_{ij}^{\mathbf{r},\mathbf{c}} = e^{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)} = \mathbf{r}(i)\mathbf{c}(j)/N \qquad (\triangleright \textit{ Fisher-Yates table})$$

Examples

- $\mu =$ Bernoulli

$$\Theta = \mathbb{R}, \quad (A, B) = (0, 1), \quad \psi(\theta) = \log \frac{1 + e^{\theta}}{2}, \quad \psi'(\theta) = \frac{e^{\theta}}{1 + e^{\theta}}, \quad \phi(x) = \log \frac{x}{1 - x}.$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = x \log x + (1 - x) \log(1 - x) \qquad \triangleright \; -\textit{Entropy(Ber(x))}$$

Examples

- $\mu = \text{Bernoulli}$

$$\Theta = \mathbb{R}, \quad (A,B) = (0,1), \quad \psi(\theta) = \log \frac{1+e^\theta}{2}, \quad \psi'(\theta) = \frac{e^\theta}{1+e^\theta}, \quad \phi(x) = \log \frac{x}{1-x}.$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = x \log x + (1-x)\log(1-x) \qquad \triangleright \textit{ $-$Entropy(Ber(x))}$$

- $\mu = \text{Counting}(\mathbb{Z}_{\geq 0})$

$$\Theta = (-\infty, 0), \quad \psi(\theta) = -\log(1-e^\theta), \quad \psi'(\theta) = \frac{e^\theta}{1-e^\theta}, \quad \phi(x) = -\log(1+x^{-1})$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = x\log x - (1+x)\log(1+x) \qquad \triangleright \textit{ $-$Entropy(Geom(x))}$$

$$Z_{ij}^{\mathbf{r},\mathbf{c}} = \frac{1}{\exp(-\boldsymbol{\alpha}(i) - \boldsymbol{\beta}(j)) - 1} \qquad \text{s.t. } Z^{\mathbf{r},\mathbf{c}} \in \mathcal{T}(\mathbf{r},\mathbf{c})$$

- $\mu = \text{Lebesgue}(\mathbb{R}_{\geq 0})$

$$\Theta = (-\infty, 0), \quad \psi(\theta) = -\log(-\theta), \quad \psi'(\theta) = -\frac{1}{\theta}, \quad \phi(x) = -\frac{1}{x}$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = -1 - \log x$$

$$Z_{ij}^{\mathbf{r},\mathbf{c}} = \frac{-1}{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)} \qquad \text{s.t. } Z^{\mathbf{r},\mathbf{c}} \in \mathcal{T}(\mathbf{r}, \mathbf{c})$$

- $\mu = \text{Lebesgue}(\mathbb{R}_{\geq 0})$

$$\Theta = (-\infty, 0), \quad \psi(\theta) = -\log(-\theta), \quad \psi'(\theta) = -\frac{1}{\theta}, \quad \phi(x) = -\frac{1}{x}$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = -1 - \log x$$

$$Z^{\mathbf{r},\mathbf{c}}_{ij} = \frac{-1}{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)} \qquad \text{s.t. } Z^{\mathbf{r},\mathbf{c}} \in \mathcal{T}(\mathbf{r}, \mathbf{c})$$

- $\mu = \text{Gamma } (\mu(dx) = x^{\gamma-1}\, dx)$

$$\Theta = (-\infty, 0), \quad \psi(\theta) = \log \Gamma(\gamma) - \gamma \log(-\theta), \quad \psi'(\theta) = -\frac{\gamma}{\theta}, \quad \phi(x) = -\frac{\gamma}{x}$$

$$f(x) = x\phi(x) - \psi(\phi(x)) = -\gamma - \gamma \log x$$

$$Z^{\mathbf{r},\mathbf{c}}_{ij} = \frac{-\gamma}{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))} \qquad \text{s.t. } Z^{\mathbf{r},\mathbf{c}} \in \mathcal{T}(\mathbf{r}, \mathbf{c})$$

- (*Informal result I: Minimum relative entropy perspective*)

  $X \sim \mu^{\otimes(m\times n)}$ conditioned on being in $\mathcal{T}(\mathbf{r},\mathbf{c})$ concentrates around $Z^{\mathbf{r},\mathbf{c}}$, where $Z_{ij}^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))$ for some $\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n$

- (*Informal result I: Minimum relative entropy perspective*)

  $X \sim \mu^{\otimes(m\times n)}$ **conditioned on being in** $\mathcal{T}(\mathbf{r},\mathbf{c})$ **concentrates around** $Z^{\mathbf{r},\mathbf{c}}$**,**
  where $Z^{\mathbf{r},\mathbf{c}}_{ij} = \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))$ **for some** $\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n$

- (*Informal result II*: Maximum likelihood perspective)

  $\Big[\, X \sim \mu^{\otimes(m\times n)}$ **conditioned on being in** $\mathcal{T}(\mathbf{r},\mathbf{c})\, \Big] \approx Y,$
  where Y **has independent entries** $Y_{ij} \sim \mu_{\alpha(i)+\boldsymbol{\beta}(j)}$ **and** $\mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}}$

- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0, 1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$

- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0,1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\,dx = \int_0^1 \mathbf{c}(y)\,dy$
- For a $m \times n$ discrete margin $(\mathbf{r}_m, \mathbf{c}_n)$, define the corresponding **continuum step margin** $(\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)$ as

$$\bar{\mathbf{r}}_m(t) := n^{-1}\mathbf{r}_m(\lceil mt \rceil), \qquad \bar{\mathbf{c}}_n(t) := m^{-1}\mathbf{c}_n(\lceil nt \rceil).$$

- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0, 1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\,dx = \int_0^1 \mathbf{c}(y)\,dy$
- For a $m \times n$ discrete margin $(\mathbf{r}_m, \mathbf{c}_n)$, define the corresponding **continuum step margin** $(\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)$ as

$$\bar{\mathbf{r}}_m(t) := n^{-1}\mathbf{r}_m(\lceil mt \rceil), \qquad \bar{\mathbf{c}}_n(t) := m^{-1}\mathbf{c}_n(\lceil nt \rceil).$$

- For $A \in \mathbb{R}^{m\times n}$, W_A := corresponding **step kernel** on unit square:

$$W_A(x, y) := A_{ij} \text{ if } (x, y) \in \left(\frac{i-1}{m}, \frac{i}{m}\right] \times \left(\frac{j-1}{n}, \frac{j}{n}\right]$$

Limit of a random matrix with i.i.d. entries given margins

- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0, 1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$
- For a $m \times n$ discrete margin $(\mathbf{r}_m, \mathbf{c}_n)$, define the corresponding **continuum step margin** $(\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)$ as

$$\bar{\mathbf{r}}_m(t) := n^{-1}\mathbf{r}_m(\lceil mt \rceil), \qquad \bar{\mathbf{c}}_n(t) := m^{-1}\mathbf{c}_n(\lceil nt \rceil).$$

- For $A \in \mathbb{R}^{m \times n}$, $W_A :=$ corresponding **step kernel** on unit square:

$$W_A(x, y) := A_{ij} \text{ if } (x, y) \in \left(\frac{i-1}{m}, \frac{i}{m}\right] \times \left(\frac{j-1}{n}, \frac{j}{n}\right]$$

- A seq. of $m \times n$ margins $(\mathbf{r}_m, \mathbf{c}_n)$ **converges in L^1** to a continuum margin $(\mathbf{r}, \mathbf{c})$ if

$$\lim_{m,n\to\infty} \|\mathbf{r} - \bar{\mathbf{r}}_m\|_1 + \|\mathbf{c} - \bar{\mathbf{c}}_n\|_1 = 0.$$

Limit of a random matrix with i.i.d. entries given margins

- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0, 1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$
- For a $m \times n$ discrete margin $(\mathbf{r}_m, \mathbf{c}_n)$, define the corresponding **continuum step margin** $(\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)$ as

$$\bar{\mathbf{r}}_m(t) := n^{-1}\mathbf{r}_m(\lceil mt \rceil), \qquad \bar{\mathbf{c}}_n(t) := m^{-1}\mathbf{c}_n(\lceil nt \rceil).$$

- For $A \in \mathbb{R}^{m\times n}$, W_A := corresponding **step kernel** on unit square:

$$W_A(x, y) := A_{ij} \text{ if } (x, y) \in \left(\frac{i-1}{m}, \frac{i}{m}\right] \times \left(\frac{j-1}{n}, \frac{j}{n}\right]$$

- A seq. of $m \times n$ margins $(\mathbf{r}_m, \mathbf{c}_n)$ **converges in L^1** to a continuum margin $(\mathbf{r}, \mathbf{c})$ if

$$\lim_{m,n\to\infty} \|\mathbf{r} - \bar{\mathbf{r}}_m\|_1 + \|\mathbf{c} - \bar{\mathbf{c}}_n\|_1 = 0.$$

- (*Informal result III*)

For $(\mathbf{r}_m, \mathbf{c}_n) \to (\mathbf{r}, \mathbf{c})$ in L^1 and $X \sim \mu^{\otimes(m\times n)}$ conditioned on $\mathcal{T}(\mathbf{r}_m, \mathbf{c}_n)$, $W_X \to W^{\mathbf{r},\mathbf{c}}$ w.h.p. in 'cut norm'

where $W^{\mathbf{r},\mathbf{c}}(x, y) = \psi'(\alpha(x) + \beta(y))$ for some $\alpha, \beta \in [0, 1] \to \mathbb{R}$.

Plan of the talk

1. Connection to Contingency tables and Phase transition
2. Connection to Relative entropy minimization and Schrödinger bridge
3. Connection to Random graphs with given degree sequence
4. Dual formulation and generalized Sinkhorn algorithm
5. Formal statement of results and some Key ideas
6. Open problems

Uniform contingency tables in statistics

- Contingency tables = matrices with non-netative integer entries with fixed row an column 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
		$X = (X_{ij})$				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

Conjecture (Independence heuristic, Good '50)

$$|\mathcal{T}(\mathbf{r},\mathbf{c})| \approx \mathrm{G}(\mathbf{r},\mathbf{c})$$

where

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

Conjecture (Independence heuristic, Good '50)

$$|\mathcal{T}(\mathbf{r},\mathbf{c})| \approx \mathrm{G}(\mathbf{r},\mathbf{c})$$

where

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

Good says: "A random table with total sum N independently satisfies the row and column margins"

Conjecture (Independence heuristic, Good '50)

$$|\mathcal{T}(\mathbf{r},\mathbf{c})| \approx \mathrm{G}(\mathbf{r},\mathbf{c})$$

where

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

Good says: "A random table with total sum N independently satisfies the row and column margins"

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

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

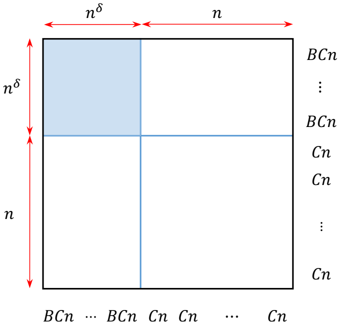
- Given implicitly by Good in 1963 [14] and later formally in 1963 [12] and 1976 [13]
- Experimentally verified by Good and Crook [11] in 1977 and Diagonis and Gangolli [8] in 1995
- Canfield and McKay '10 [5]: For $m = n$ and $\mathbf{r} = \mathbf{c} = (\lfloor Cn \rfloor, \dots, \lfloor Cn \rfloor)$,

$$\begin{aligned}\log|\mathcal{T}(\mathbf{r},\mathbf{c})| &= [(1+C)\log(1+C) - C\log(C)]n^2 - n\log n \\ &\qquad - n\log 2\pi C(1+C) + \log n + O(1) \\ &\sim \log\sqrt{e}\,\mathrm{G}(\mathbf{r},\mathbf{c})\end{aligned}$$

- In 2008, Greenhill and McKay [15] proved same asymptotics for uniform but sparse margins: $\max(\mathbf{r}) \cdot \max(\mathbf{c}) = O(N^{2/3})$

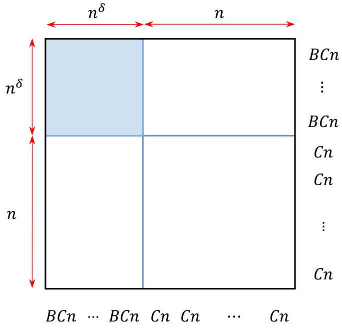
But what about non-uniform margins?

- 2×2 block (Barvinok) margins: $\mathbf{r} = \mathbf{c} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{(n-n^\delta)})$, $0 \le \delta \le 1$



But what about non-uniform margins?

- 2×2 block (Barvinok) margins: $\mathbf{r} = \mathbf{c} = (\overbrace{BCn, \ldots, BCn}^{n^\delta}, \overbrace{Cn, \ldots\ldots, Cn}^{(n-n^\delta)})$, $0 \le \delta \le 1$



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

$$\lim_{n\to\infty} \frac{1}{n^2} \log |\mathcal{T}(\mathbf{r}, \mathbf{c})| > \lim_{n\to\infty} \frac{1}{n^2} \log \mathrm{G}(\mathbf{r}, \mathbf{c}).$$

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

Typical table

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* $\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c}))$"

Typical table

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* Uniform($\mathcal{T}(\mathbf{r},\mathbf{c})$)"
- (Barvinok '10 [3]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \arg\max_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{\sum_{i,j} (x_{ij}+1)\log(x_{ij}+1) - x_{ij}\log(x_{ij})}_{=g(X)}$$

$$= \arg\max_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} \text{Entropy}(\text{Geom}(\text{mean} = x_{ij}))$$

Typical table

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* $\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c}))$"
- (Barvinok '10 [3]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,max}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{\sum_{i,j} (x_{ij}+1)\log(x_{ij}+1) - x_{ij}\log(x_{ij})}_{=g(X)}$$

$$= \operatorname*{arg\,max}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} \text{Entropy}(\text{Geom}(\text{mean} = x_{ij}))$$

- Barvinok's insight:

$$\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c})) \approx Z^{\mathbf{r},\mathbf{c}}$$

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* Uniform($\mathcal{T}(\mathbf{r},\mathbf{c})$)"
- (Barvinok '10 [3]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,max}_{X\in\mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{\sum_{i,j} (x_{ij}+1)\log(x_{ij}+1) - x_{ij}\log(x_{ij})}_{=g(X)}$$

$$= \operatorname*{arg\,max}_{X\in\mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} \text{Entropy}(\text{Geom}(\text{mean} = x_{ij}))$$

- Barvinok's insight:

$$\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c})) \approx Z^{\mathbf{r},\mathbf{c}}$$

  - (Barvinok '09 [1], '10 [3])

$$g(Z^{\mathbf{r},\mathbf{c}}) - \gamma n \log n \le \log|\mathcal{T}(\mathbf{r},\mathbf{c})| \le g(Z^{\mathbf{r},\mathbf{c}})$$

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* Uniform$(\mathcal{T}(\mathbf{r},\mathbf{c}))$"
- (Barvinok '10 [3]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \arg\max_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{\sum_{i,j} (x_{ij}+1)\log(x_{ij}+1) - x_{ij}\log(x_{ij})}_{=g(X)}$$

$$= \arg\max_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} \text{Entropy}(\text{Geom}(\text{mean} = x_{ij}))$$

- Barvinok's insight:

$$\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c})) \approx Z^{\mathbf{r},\mathbf{c}}$$

  - (Barvinok '09 [1], '10 [3])

$$g(Z^{\mathbf{r},\mathbf{c}}) - \gamma n \log n \leq \log|\mathcal{T}(\mathbf{r},\mathbf{c})| \leq g(Z^{\mathbf{r},\mathbf{c}})$$

  - Brändén, Leake, and Pak '23 [4] generalized this result to CTs with possibly bounded integer values (Using Lorenzian polynomials)

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* Uniform$(\mathcal{T}(\mathbf{r},\mathbf{c}))$"
- (Barvinok '10 [3]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \arg\max_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{\sum_{i,j} (x_{ij}+1)\log(x_{ij}+1) - x_{ij}\log(x_{ij})}_{=g(X)}$$

$$= \arg\max_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} \text{Entropy}(\text{Geom}(\text{mean} = x_{ij}))$$

$$= \arg\min_{X \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} D\left(\mu_{\phi(x_{ij})} \,\|\, \mu = \text{Counting}(\mathbb{Z}_{\geq 0})\right)$$

- Barvinok's insight:

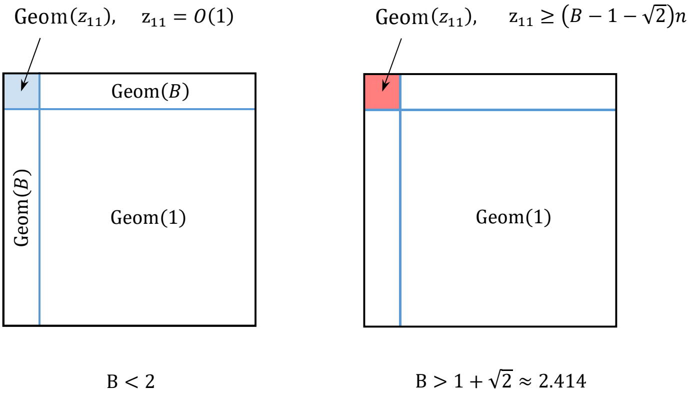
$$\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c})) \approx Z^{\mathbf{r},\mathbf{c}}$$

  - (Barvinok '09 [1], '10 [3])

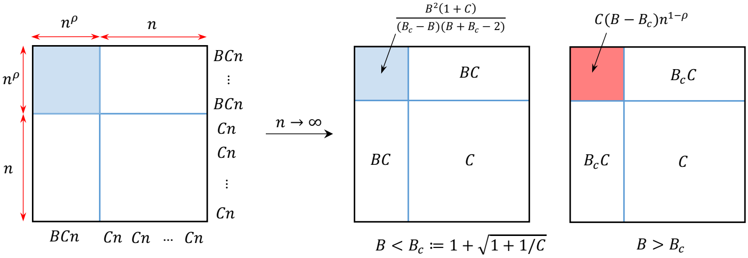
$$g(Z^{\mathbf{r},\mathbf{c}}) - \gamma n \log n \leq \log|\mathcal{T}(\mathbf{r},\mathbf{c})| \leq g(Z^{\mathbf{r},\mathbf{c}})$$

  - Brändén, Leake, and Pak '23 [4] generalized this result to CTs with possibly bounded integer values (Using Lorenzian polynomials)

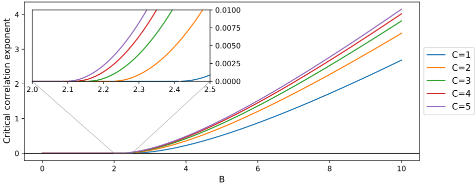
- In 2010, Barbinok conjectured that there is a phase transition in Uniform($\mathcal{T}$(Barv. margin)) as B increases



- Typical tables can change drastically by a small change in the margin!
  - For $0 \leq \delta < 1$, Dittmer, Lyu, and Pak [9] show that $Z^{\mathbf{r},\mathbf{c}}$ undergoes a **sharp phase transition** at $B_c = 1 + \sqrt{1 + C^{-1}}$:



  - This result was used to obtain a second-order phase transition in the number of CTs with Barvinok margin by Lyu and Pak '22 [17]



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

- Given a base probability measure $\mathcal{R}$ on $\mathbb{R}^2$ and marginal distributions μ_1 and μ_2,

$$(**) \qquad \min_{\mathcal{H}\in\Pi(\mu_1,\mu_2)} D_{KL}(\mathcal{H}\,\|\,\mathcal{R})$$

The optimal $\mathcal{H}$ from above is the **static Schrödinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$ [10, 18]

- Given a base probability measure $\mathcal{R}$ on $\mathbb{R}^2$ and marginal distributions μ_1 and μ_2,

$$(**) \qquad \min_{\mathcal{H} \in \Pi(\mu_1, \mu_2)} D_{KL}(\mathcal{H} \,\|\, \mathcal{R})$$

  The optimal $\mathcal{H}$ from above is the **static Schrödinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$ [10, 18]

- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [19] s.t.

$$\frac{d\mathcal{H}}{d\mathcal{R}}(x, y) = e^{\boldsymbol{\alpha}_1(x) + \boldsymbol{\alpha}_2(y)} \qquad \mathcal{R}\text{-a.s.}$$

- Given a base probability measure $\mathcal{R}$ on $\mathbb{R}^2$ and marginal distributions μ_1 and μ_2,

$$(**) \qquad \min_{\mathcal{H} \in \Pi(\mu_1,\mu_2)} D_{KL}(\mathcal{H} \,\|\, \mathcal{R})$$

  The optimal $\mathcal{H}$ from above is the **static Schrödinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$ [10, 18]
- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [19] s.t.

$$\frac{d\mathcal{H}}{d\mathcal{R}}(x, y) = e^{\boldsymbol{\alpha}_1(x)+\boldsymbol{\alpha}_2(y)} \qquad \mathcal{R}\text{-a.s.}$$

- Specializing $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, $(**)$ becomes **Entropic Optimal Transport**:

$$\min_{\mathcal{H} \in \Pi(\mu_1,\mu_2)} \int_{\mathbb{R}^2} \gamma(x, y)\, \mathcal{H}(dx, dy) + \varepsilon D_{KL}(\mathcal{H} \,\|\, \mu_1 \otimes \mu_2),$$

- Given a base probability measure $\mathcal{R}$ on $\mathbb{R}^2$ and marginal distributions μ_1 and μ_2,

$$(**) \qquad \min_{\mathcal{H}\in\Pi(\mu_1,\mu_2)} D_{KL}(\mathcal{H}\,\|\,\mathcal{R})$$

  The optimal $\mathcal{H}$ from above is the **static Schrödinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$ [10, 18]

- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [19] s.t.

$$\frac{d\mathcal{H}}{d\mathcal{R}}(x,y) = e^{\boldsymbol{\alpha}_1(x)+\boldsymbol{\alpha}_2(y)} \qquad \mathcal{R}\text{-a.s.}$$

- Specializing $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, $(**)$ becomes **Entropic Optimal Transport**:

$$\min_{\mathcal{H}\in\Pi(\mu_1,\mu_2)} \int_{\mathbb{R}^2} \gamma(x,y)\,\mathcal{H}(dx,dy) + \varepsilon D_{KL}(\mathcal{H}\,\|\,\mu_1\otimes\mu_2),$$

- Specializing $\mathcal{R} = \text{Uniform}([m]\times[n])$, $(**)$ becomes

$$\min_{X=(x_{ij})\in(0,\infty)^{m\times n}} x_{ij}\log x_{ij} \qquad \text{subject to} \quad \sum_{j=1}^{n} x_{ij} = \mu_1(i), \quad \sum_{i=1}^{m} x_{ij} = \mu_2(j)\ \forall i,j.$$

- Given a base probability measure $\mathcal{R}$ on $\mathbb{R}^2$ and marginal distributions μ_1 and μ_2,

$$(**) \qquad \min_{\mathcal{H}\in\Pi(\mu_1,\mu_2)} D_{KL}(\mathcal{H}\,\|\,\mathcal{R})$$

The optimal $\mathcal{H}$ from above is the **static Schrödinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$ [10, 18]

- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [19] s.t.

$$\frac{d\mathcal{H}}{d\mathcal{R}}(x,y) = e^{\boldsymbol{\alpha}_1(x)+\boldsymbol{\alpha}_2(y)} \qquad \mathcal{R}\text{-a.s.}$$

- Specializing $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, $(**)$ becomes **Entropic Optimal Transport**:

$$\min_{\mathcal{H}\in\Pi(\mu_1,\mu_2)} \int_{\mathbb{R}^2} \gamma(x,y)\,\mathcal{H}(dx,dy) + \varepsilon D_{KL}(\mathcal{H}\,\|\,\mu_1\otimes\mu_2),$$

- Specializing $\mathcal{R} = \text{Uniform}([m]\times[n])$, $(**)$ becomes

$$\min_{X=(x_{ij})\in(0,\infty)^{m\times n}} x_{ij}\log x_{ij} \qquad \text{subject to} \quad \sum_{j=1}^{n} x_{ij} = \mu_1(i), \quad \sum_{i=1}^{m} x_{ij} = \mu_2(j)\ \forall i,j.$$

This is in fact the **typical table** problem with $\mu = \text{Poisson}(1)$!

- $x_{ij}\log x_{ij} = D(\mu_{\phi(x_{ij})}\,\|\,\mu) + x_{ij} - 1$

- $\mathbf{d} = (d_1, d_2, \ldots, d_n)$: **degree sequence** of an n-node graph

- $\mathbf{d} = (d_1, d_2, \ldots, d_n)$: **degree sequence** of an n-node graph
- (*Question*)

How does a uniformly random graph with degree sequence d look like?

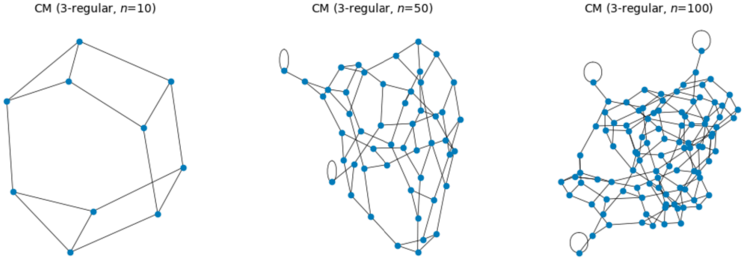


Figure: Random 3-regular graphs generated by the configuration model (allowing loops)

- $(\mathbf{d}^n)_{n\geq 1}$: dense degree sequence with scaling limit to $\mathbf{c} : [0,1] \to (c_1, c_2) \subseteq (0,1)$

Limit of random graphs with given degree sequences

- $(\mathbf{d}^n)_{n\geq 1}$: dense degree sequence with scaling limit to $\mathbf{c} : [0,1] \to (c_1, c_2) \subseteq (0,1)$
- Chaterjee, Diaconis, Sly '11 [6]

- $(\mathbf{d}^n)_{n\geq 1}$: dense degree sequence with scaling limit to $\mathbf{c} : [0,1] \to (c_1, c_2) \subseteq (0,1)$
- Chaterjee, Diaconis, Sly '11 [6]
  - Assume $\mathbf{c}$ satisfies the 'continuum Erdős-Gallai condition'

$$\int_x^1 \mathbf{c}(y) \wedge x \, dy + x^2 - \int_0^x \mathbf{c}(y) \, dy > 0$$

Limit of random graphs with given degree sequences

- $(\mathbf{d}^n)_{n\geq 1}$: dense degree sequence with scaling limit to $\mathbf{c} : [0,1] \to (c_1, c_2) \subseteq (0,1)$
- Chaterjee, Diaconis, Sly '11 [6]
  - Assume $\mathbf{c}$ satisfies the 'continuum Erdős-Gallai condition'

$$\int_x^1 \mathbf{c}(y) \wedge x\, dy + x^2 - \int_0^x \mathbf{c}(y)\, dy > 0$$

  - Then there exists a limiting 'continuum dual variable ' $\boldsymbol{\beta}^* : [0,1] \to \mathbb{R}$ such that the corresponding graphon

$$W^{\boldsymbol{\beta}^*}(x,y) = \frac{1}{\exp(\boldsymbol{\beta}^*(x) + \boldsymbol{\beta}^*(y)) + 1}$$

has 'degree sequence' $\mathbf{c}$:

$$\int_{\mathbb{R}} W^{\boldsymbol{\beta}^*}(x,y)\, dy = \mathbf{c}(x)$$

Limit of random graphs with given degree sequences

- $(\mathbf{d}^n)_{n\geq 1}$: dense degree sequence with scaling limit to $\mathbf{c} : [0,1] \to (c_1, c_2) \subseteq (0,1)$
- Chaterjee, Diaconis, Sly '11 [6]
  - Assume $\mathbf{c}$ satisfies the 'continuum Erdős-Gallai condition'

$$\int_x^1 \mathbf{c}(y) \wedge x \, dy + x^2 - \int_0^x \mathbf{c}(y) \, dy > 0$$

  - Then there exists a limiting 'continuum dual variable ' $\boldsymbol{\beta}^* : [0,1] \to \mathbb{R}$ such that the corresponding graphon

$$W^{\boldsymbol{\beta}^*}(x,y) = \frac{1}{\exp(\boldsymbol{\beta}^*(x) + \boldsymbol{\beta}^*(y)) + 1}$$

has 'degree sequence' $\mathbf{c}$:

$$\int_{\mathbb{R}} W^{\boldsymbol{\beta}^*}(x,y) \, dy = \mathbf{c}(x)$$

  - A^n = Adj mx of the uniformly random graph with degree seq. $\mathbf{d}^n$. Then

$$W_{A^n} \to W^{\boldsymbol{\beta}^*} \qquad \textit{in weak cut distance},$$

(W_{A^n}: step function corresponding to the adj mx A^n)

Key idea in CDS '11 [6]

- Fit a parametrized random graph model with independent edges (s.t. deg seq = sufficient statistic) to the target degree sequence by MLE

Key idea in CDS '11 [6]

- Fit a parametrized random graph model with independent edges (s.t. deg seq = sufficient statistic) to the target degree sequence by MLE
  - (The β-model) Given a dual variable $\boldsymbol{\beta} : \mathbb{R}^n$, $G_{\boldsymbol{\beta}}$ =random graph with n nodes and independent edges, where

$$\mathbb{P}_{\boldsymbol{\beta}}(i \sim j) \propto \exp(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j))$$
$$= \text{exponetial tilting of Uniform}(\{0,1\}) \text{ by } \boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)$$

Expected adjacency matrix:

$$\mathbb{E}[A^{\boldsymbol{\beta}}(i,j)] := \frac{e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}}{1 + e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}} = \psi'(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)),$$

- Fit a parametrized random graph model with independent edges (s.t. deg seq = sufficient statistic) to the target degree sequence by MLE
  - (The β-model) Given a dual variable $\boldsymbol{\beta} : \mathbb{R}^n$, $G_{\boldsymbol{\beta}}$ =random graph with n nodes and independent edges, where

$$\mathbb{P}_{\boldsymbol{\beta}}(i \sim j) \propto \exp(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j))$$

$$= \text{exponetial tilting of Uniform}(\{0,1\}) \text{ by } \boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)$$

Expected adjacency matrix:

$$\mathbb{E}[A^{\boldsymbol{\beta}}(i,j)] := \frac{e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}}{1 + e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}} = \psi'(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)),$$

  - (log-likelihood)

$$\ell(\boldsymbol{\beta}) = \sum_{i,j} x_{ij}(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)) - \psi(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)) = 2\langle \mathbf{d}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j))$$

Key idea in CDS '11 [6]

- Fit a parametrized random graph model with independent edges (s.t. deg seq = sufficient statistic) to the target degree sequence by MLE
  - (The β-model) Given a dual variable $\boldsymbol{\beta} : \mathbb{R}^n$, $G_{\boldsymbol{\beta}}$ =random graph with n nodes and independent edges, where

$$\mathbb{P}_{\boldsymbol{\beta}}(i \sim j) \propto \exp(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j))$$
$$= \text{exponetial tilting of Uniform}(\{0,1\}) \text{ by } \boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)$$

Expected adjacency matrix:

$$\mathbb{E}[A^{\boldsymbol{\beta}}(i,j)] := \frac{e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}}{1 + e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}} = \psi'(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)),$$

  - (log-likelihood)

$$\ell(\boldsymbol{\beta}) = \sum_{i,j} x_{ij}(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)) - \psi(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)) = 2\langle \mathbf{d}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j))$$

  - (The MLE equation) $\frac{d\ell(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = 0 \iff \mathbb{E}[\text{degree seq.}] = \mathbf{d}$:

$$\mathbb{E}\left[\sum_{j=1}^{n} A^{\boldsymbol{\beta}}(i,j)\right] = d_i \qquad \text{for all } 1 \le i \le n$$

- **Sketch of proof**:
  - Find MLE $\boldsymbol{\beta}^n$ for the $\boldsymbol{\beta}$-model to the target degree sequence $\mathbf{d}^n$
  - Show that the MLEs $\boldsymbol{\beta}^n$ converge (after scaling) to some $\boldsymbol{\beta}^* : [0,1] \to \mathbb{R}$ in L^1
  - Show that the expected adjacency matrices of the ML $\boldsymbol{\beta}$-model converges to the limiting graphon:

$$W_{\mathbb{E}[A^{\boldsymbol{\beta}^n}]} \to W^{\boldsymbol{\beta}^*}$$

  - Show that the $\boldsymbol{\beta}^n$-model concentrates around its mean (in weak cut distance)

$$W_{A^{\boldsymbol{\beta}_n}} \approx W_{\mathbb{E}[A^{\boldsymbol{\beta}_n}]}$$

  - Show that the $\boldsymbol{\beta}^n$-model as the target deg. seq. $\mathbf{d}^n$ with prob. $\geq \exp(-o(n^{2/3+\varepsilon}))$
  - Putting things together:

$$W_{A^{\boldsymbol{\beta}_n}} \overset{\text{weak cut}}{\approx} W_{\mathbb{E}[A^{\boldsymbol{\beta}^n}]} = W^{\boldsymbol{\beta}^*} + o(1)$$

1. Sharp sufficient conditions for the ‘subcritical regime’
2. Concentration of margin-constrained random matrices around the typical table
3. Cut-norm scailing limit of a sequence of margin-constrained random matrices to the typical kernel
4. (Optional) Linear convergence of generalized Sinkhorn algorithm

- Back to the margin-conditioned random matrix (RM) setting.

- Back to the margin-conditioned random matrix (RM) setting.
- Recall the (generalized) typical table

$$Z^{\mathbf{r},\mathbf{c}} = \operatorname*{arg\,min}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m\times n}} \sum_{i,j} D(\mu_{\phi(x_{ij})} \,\|\, \mu)$$

Dual(Typical Table)=MLE

- Back to the margin-conditioned random matrix (RM) setting.
- Recall the (generalized) typical table

$$Z^{\mathbf{r},\mathbf{c}} = \operatorname*{arg\,min}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m \times n}} \sum_{i,j} D(\mu_{\phi(x_{ij})} \,\|\, \mu)$$

- A natural statistical model for margin-conditioned RM $=$ $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$: independent entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}$
  - MLE $\iff \mathbb{E}[Y] \in \mathcal{T}(\mathbf{r}, \mathbf{c})$

Dual(Typical Table)=MLE

- Back to the margin-conditioned random matrix (RM) setting.
- Recall the (generalized) typical table

$$Z^{\mathbf{r},\mathbf{c}} = \operatorname*{arg\,min}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m \times n}} \sum_{i,j} D(\mu_{\phi(x_{ij})} \,\|\, \mu)$$

- A natural statistical model for margin-conditioned RM $=$ $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$: independent entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}$
  - MLE $\iff \mathbb{E}[Y] \in \mathcal{T}(\mathbf{r}, \mathbf{c})$
- (**Strong duality**; this work [16])

- Back to the margin-conditioned random matrix (RM) setting.
- Recall the (generalized) typical table

$$Z^{\mathbf{r},\mathbf{c}} = \operatorname*{arg\,min}_{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m\times n}} \sum_{i,j} D(\mu_{\phi(x_{ij})} \,\|\, \mu)$$

- A natural statistical model for margin-conditioned RM = $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: independent entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$
  - MLE $\Longleftrightarrow \mathbb{E}[Y] \in \mathcal{T}(\mathbf{r}, \mathbf{c})$
- (**Strong duality**; this work [16])
  - $$\operatorname*{arg\,min}_{Z} \underline{\texttt{Relative Entropy}}(Z) = \operatorname*{arg\,max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \underline{\texttt{Likelihood Of Margin}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$
    $$= \operatorname*{arg\,max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(\langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$

- Back to the margin-conditioned random matrix (RM) setting.
- Recall the (generalized) typical table

$$Z^{\mathbf{r},\mathbf{c}} = \underset{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m\times n}}{\arg\min} \sum_{i,j} D(\mu_{\phi(x_{ij})} \,\|\, \mu)$$

- A natural statistical model for margin-conditioned RM = $(\boldsymbol{\alpha},\boldsymbol{\beta})$-model
  - $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: independent entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$
  - MLE $\iff \mathbb{E}[Y] \in \mathcal{T}(\mathbf{r},\mathbf{c})$
- (**Strong duality**; this work [16])
  - $$\underset{Z}{\arg\min}\ \underline{\texttt{Relative Entropy}}(Z) = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\arg\max}\ \underline{\texttt{Likelihood Of Margin}}(\boldsymbol{\alpha},\boldsymbol{\beta})$$
    $$= \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\arg\max} \left(\langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)) \right)$$
  - For $\psi' = \texttt{tilt2mean}$ function,

$$\underbrace{Z^{\mathbf{r},\mathbf{c}}_{ij}}_{\text{typical table}} = \psi'\Big(\underbrace{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}_{=\text{MLE}} \Big)$$

Dual(Typical Table)=MLE

- Back to the margin-conditioned random matrix (RM) setting.
- Recall the (generalized) typical table

$$Z^{\mathbf{r},\mathbf{c}} = \underset{X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \cap (A,B)^{m\times n}}{\arg\min} \sum_{i,j} D(\mu_{\phi(x_{ij})} \,\|\, \mu)$$

- A natural statistical model for margin-conditioned RM = $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: independent entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$
  - MLE $\Longleftrightarrow \mathbb{E}[Y] \in \mathcal{T}(\mathbf{r}, \mathbf{c})$
- (**Strong duality**; this work [16])
  - $$\arg\min_{Z} \underline{\texttt{Relative Entropy}}(Z) = \arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \underline{\texttt{Likelihood Of Margin}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$
    $$= \arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(\langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$
  - For $\psi' = \texttt{tilt2mean}$ function,
    $$\underbrace{Z_{ij}^{\mathbf{r},\mathbf{c}}}_{\text{typical table}} = \psi'\Big(\underbrace{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}_{=\text{MLE}} \Big)$$
  - How do we compute MLE??

Tame margins

- When does the typical table/MLE exists?

$$\underbrace{Z^{\mathbf{r},\mathbf{c}}_{ij}}_{\text{typical table}} = \psi'\Big(\underbrace{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}_{=\text{MLE}}\Big)$$

Furthermore, when is the MLE uniformly bounded? ('Subcritical Regime')

- When does the typical table/MLE exists?

$$\underbrace{Z_{ij}^{\mathbf{r},\mathbf{c}}}_{\text{typical table}} = \psi'\Big(\underbrace{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}_{=\text{MLE}}\Big)$$

Furthermore, when is the MLE uniformly bounded? ('Subcritical Regime')

- (Tame margins) For $\delta > 0$, a margin $(\mathbf{r}, \mathbf{c})$ is δ-**tame** if the typical table $Z^{\mathbf{r},\mathbf{c}}$ exists and its entries satisfy

$$A < A_\delta := \max\left(A + \delta, -\delta^{-1}\right) \leq Z^{\mathbf{r},\mathbf{c}} \leq \min\left(B - \delta, \delta^{-1}\right) =: B_\delta < B$$

- When does the typical table/MLE exists?

$$\underbrace{Z_{ij}^{\mathbf{r},\mathbf{c}}}_{\text{typical table}} = \psi'\Big(\underbrace{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}_{=\text{MLE}}\Big)$$

Furthermore, when is the MLE uniformly bounded? ('Subcritical Regime')

- (Tame margins) For $\delta > 0$, a margin $(\mathbf{r}, \mathbf{c})$ is δ-**tame** if the typical table $Z^{\mathbf{r},\mathbf{c}}$ exists and its entries satisfy

$$A < A_\delta := \max\left(A + \delta, -\delta^{-1}\right) \leq Z^{\mathbf{r},\mathbf{c}} \leq \min\left(B - \delta, \delta^{-1}\right) =: B_\delta < B$$

  - For $\mu = \text{Counting}(\mathbb{Z}_{\geq 0})$, Barvinok margin $\mathbf{r} = \mathbf{c} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \ldots, \lfloor Cn \rfloor)$, Dittmer, L., Pak '20 [9]:

$$\text{tame} \quad \Longleftrightarrow \quad B < B_c = 1 + \sqrt{1 + C^{-1}}$$

- When does the typical table/MLE exists?

$$\underbrace{Z^{\mathbf{r},\mathbf{c}}_{ij}}_{\text{typical table}} = \psi'\Big(\underbrace{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}_{=\text{MLE}}\Big)$$

Furthermore, when is the MLE uniformly bounded? ('Subcritical Regime')

- (Tame margins) For $\delta > 0$, a margin $(\mathbf{r},\mathbf{c})$ is δ-**tame** if the typical table $Z^{\mathbf{r},\mathbf{c}}$ exists and its entries satisfy

$$A < A_\delta := \max\left(A+\delta, -\delta^{-1}\right) \le Z^{\mathbf{r},\mathbf{c}} \le \min\left(B-\delta, \delta^{-1}\right) =: B_\delta < B$$

  - For $\mu = \text{Counting}(\mathbb{Z}_{\ge 0})$, Barvinok margin $\mathbf{r} = \mathbf{c} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \dots, \lfloor Cn \rfloor)$, Dittmer, L., Pak '20 [9]:

$$\text{tame} \quad \Longleftrightarrow \quad B < B_c = 1 + \sqrt{1 + C^{-1}}$$

  - For $\mu = \text{Bernoulli}(1/2)$ and $sn \le \mathbf{r} = \mathbf{c} \le tn$, Barvinok and Hartigan '10 [2]:

$$\text{tame} \quad \Longleftarrow \quad (s+t)^2 < 4s$$

$$\exists \text{ non-tame margins} \quad \Longleftarrow \quad (s+t)^2 > 4s$$

Theorem (**L**-Mukherjee '24+)

If ψ'' is non-decreasing (necessarily unbounded support for μ), then for arbitrary margin $(\mathbf{r},\mathbf{c})$ with $s \leq \mathbf{r}/n$, $\mathbf{c}/m \leq t$,

$$\begin{aligned} (\mathbf{r},\mathbf{c}) \text{ is tame} &\quad \Longleftarrow \quad \phi(t) - 2\phi(s) < \phi(\infty) \\ \exists \text{ non-tame margins} &\quad \Longleftarrow \quad \phi(t) - 2\phi(s) > \phi(\infty) \end{aligned}$$

Theorem (**L**-Mukherjee '24+)

If ψ'' *is non-decreasing (necessarily unbounded support for* μ*), then for arbitrary margin* $(\mathbf{r}, \mathbf{c})$ *with* $s \leq \mathbf{r}/n,\ \mathbf{c}/m \leq t$*,*

$$\begin{aligned} (\mathbf{r}, \mathbf{c}) \text{ is tame} \quad &\Longleftarrow \quad \phi(t) - 2\phi(s) < \phi(\infty) \\ \exists \text{ non-tame margins} \quad &\Longleftarrow \quad \phi(t) - 2\phi(s) > \phi(\infty) \end{aligned}$$

In particular, for $\mu = \text{Counting}(\mathbb{Z}_{\geq 0})$*,*

$$\begin{aligned} (\mathbf{r}, \mathbf{c}) \text{ is tame} \quad &\Longleftarrow \quad t < 1 + \sqrt{1 + s^{-1}} \\ \exists \text{ non-tame margins} \quad &\Longleftarrow \quad t > 1 + \sqrt{1 + s^{-1}} \end{aligned}$$

Theorem (**L**-Mukherjee '24+)

If ψ'' is non-decreasing (necessarily unbounded support for μ), then for arbitrary margin $(\mathbf{r}, \mathbf{c})$ with $s \le \mathbf{r}/n$, $\mathbf{c}/m \le t$,

$$(\mathbf{r}, \mathbf{c}) \text{ is tame} \quad \Longleftarrow \quad \phi(t) - 2\phi(s) < \phi(\infty)$$

$$\exists \text{ non-tame margins} \quad \Longleftarrow \quad \phi(t) - 2\phi(s) > \phi(\infty)$$

In particular, for $\mu = \text{Counting}(\mathbb{Z}_{\ge 0})$,

$$(\mathbf{r}, \mathbf{c}) \text{ is tame} \quad \Longleftarrow \quad t < 1 + \sqrt{1 + s^{-1}}$$

$$\exists \text{ non-tame margins} \quad \Longleftarrow \quad t > 1 + \sqrt{1 + s^{-1}}$$

Theorem (**L**-Mukherjee '24+)

If μ has bounded support $[0, B]$, then for arbitrary margin $(\mathbf{r}, \mathbf{c})$ with $s \le \mathbf{r}/n$, $\mathbf{c}/m \le t$,

$$(\mathbf{r}, \mathbf{c}) \text{ is tame} \quad \Longleftarrow \quad (s+t)^2 < 4sB$$

$$\exists \text{ non-tame margins} \quad \Longleftarrow \quad (s+t)^2 > 4sB$$

Theorem (Stability of Schrödinger Bridge and Potentials; **L**-Mukherjee '24+)

$(\mathbf{r}_m, \mathbf{c}_n)$ = *seq. of* $m \times n$ δ*-tame margins* $\rightarrow$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1.

(i) $\exists$ *bounded measurable* $\boldsymbol{\alpha}, \boldsymbol{\beta} : [0,1] \rightarrow \mathbb{R}$ *such that the kernel*

$$W^{\mathbf{r},\mathbf{c}}(x,y) := \psi'(\boldsymbol{\alpha}(x) + \boldsymbol{\beta}(y))$$

has continuum margin $(\mathbf{r}, \mathbf{c})$.

Theorem (Stability of Schrödinger Bridge and Potentials; **L**-Mukherjee '24+)

$(\mathbf{r}_m, \mathbf{c}_n)$ = *seq. of* $m \times n$ δ*-tame margins* $\to$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1*.*

(i) $\exists$ *bounded measurable* $\boldsymbol{\alpha}, \boldsymbol{\beta} : [0,1] \to \mathbb{R}$ *such that the kernel*

$$W^{\mathbf{r},\mathbf{c}}(x,y) := \psi'(\boldsymbol{\alpha}(x) + \boldsymbol{\beta}(y))$$

has continuum margin $(\mathbf{r}, \mathbf{c})$*.*

(ii)

$$\|W^{\mathbf{r},\mathbf{c}} - W_{Z^{\mathbf{r}_m,\mathbf{c}_n}}\|_2^2 \le C_\delta \|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1$$

$$\|\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}_m\|_2^2 + \|\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}_n\|_2^2 \le C_\delta \|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1.$$

Exact conditioning for contingency tables

$$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W(x,y)\,dx\,dy\right|$$

$$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W(x,y)\,dx\,dy\right|$$

Theorem (Main result; **L**-Mukherjee '24+)

Let $(\mathbf{r}_m, \mathbf{c}_n)$ = *seq. of* $m \times n$ δ*-tame margins* $\to$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1.

$\|W\|_{\square} := \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W(x,y)\,dx\,dy\right|$

Theorem (Main result; **L**-Mukherjee '24+)

Let $(\mathbf{r}_m, \mathbf{c}_n) =$ *seq. of* $m \times n$ δ*-tame margins* $\to$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1*.*

$X \sim \mu^{\otimes(m\times n)}$ *given* $X \in \mathcal{T}(\mathbf{r}, \mathbf{c})$

$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W(x,y)\,dx\,dy \right|$

Theorem (Main result; L-Mukherjee '24+)

Let $(\mathbf{r}_m, \mathbf{c}_n)$ = *seq. of* $m \times n$ δ*-tame margins* $\to$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1*.*

$X \sim \mu^{\otimes(m\times n)}$ *given* $X \in \mathcal{T}(\mathbf{r}, \mathbf{c})$

With probability at least $1 - \exp(-(m+n)\log mn)$*,*

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \le C_1\sqrt{(m^{-1}+n^{-1})\log mn} + C_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1};$$

$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W(x,y)\,dx\,dy\right|$

Theorem (Main result; **L**-Mukherjee '24+)

Let $(\mathbf{r}_m, \mathbf{c}_n) =$ *seq. of* $m \times n$ δ*-tame margins* $\to$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1.

$X \sim \mu^{\otimes(m\times n)}$ *given* $X \in \mathcal{T}(\mathbf{r}, \mathbf{c})$

With probability at least $1 - \exp(-(m+n)\log mn)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \leq C_1\sqrt{(m^{-1}+n^{-1})\log mn} + C_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1};$$

(i) *when* $\mu =$*Counting*$([0,B]\cap\mathbb{Z})$ *or Lebesgue*$(\mathbb{R}_{\geq 0})$

$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W(x,y)\,dx\,dy\right|$

Theorem (Main result; L-Mukherjee '24+)

Let $(\mathbf{r}_m, \mathbf{c}_n) =$ *seq. of* $m \times n$ δ*-tame margins* $\to$ *continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1*.*

$X \sim \mu^{\otimes(m\times n)}$ *given* $X \in \mathcal{T}(\mathbf{r}, \mathbf{c})$

With probability at least $1 - \exp(-(m+n)\log mn)$*,*

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \leq C_1\sqrt{(m^{-1}+n^{-1})\log mn} + C_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1};$$

(i) *when* $\mu =$*Counting*$([0,B]\cap\mathbb{Z})$ *or Lebesgue*$(\mathbb{R}_{\geq 0})$

(ii) *If* μ *admits a positive density w.r.t. the measure in* ***(i)****, similar result holds but with slightly worse rate.*

Concentration of large random matrices with given margin

$$\mathcal{T}_\rho(\mathbf{r},\mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m\times n} \,:\, \|r(\mathbf{x}) - \mathbf{r}\|_\infty \leq \rho \text{ and } \|c(\mathbf{x}) - \mathbf{c}\|_\infty \leq \rho \right\}$$

$$\mathcal{T}_\rho(\mathbf{r},\mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m\times n} \,:\, \|r(\mathbf{x}) - \mathbf{r}\|_\infty \le \rho \text{ and } \|c(\mathbf{x}) - \mathbf{c}\|_\infty \le \rho \right\}$$

Theorem (Concentration; **L**-Mukherjee '24+)

$(\mathbf{r},\mathbf{c}) = (m \times n)$ δ-tame margin, $Z = Z^{\mathbf{r},\mathbf{c}}$, $(\boldsymbol{\alpha},\boldsymbol{\beta}) = \mathrm{MLE}$, $X = (X_{ij}) \sim \mu^{\otimes(m\times n)}$. Let $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$. Then

$$\mathbb{P}\left(\|W_X - W_Z\|_\square \ge t \,\middle|\, X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \right) \le \mathbb{P}\left(Y \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right)^{-1} \exp\left(C(m+n)(1+\rho) - \frac{t^2 mn}{2C} \right).$$

$$\mathcal{T}_\rho(\mathbf{r},\mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m\times n} \,:\, \|r(\mathbf{x})-\mathbf{r}\|_\infty \le \rho \text{ and } \|c(\mathbf{x})-\mathbf{c}\|_\infty \le \rho \right\}$$

Theorem (Concentration; **L**-Mukherjee '24+)

$(\mathbf{r},\mathbf{c}) = (m\times n)$ δ-tame margin, $Z = Z^{\mathbf{r},\mathbf{c}}$, $(\boldsymbol{\alpha},\boldsymbol{\beta}) =$ MLE, $X = (X_{ij}) \sim \mu^{\otimes(m\times n)}$. Let $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$. Then

$$\mathbb{P}\left(\|W_X - W_Z\|_\square \ge t \,\middle|\, X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \right) \le \mathbb{P}\left(Y \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right)^{-1} \exp\left(C(m+n)(1+\rho) - \frac{t^2 mn}{2C} \right).$$

Lemma

Let $Y \sim \mu_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$ be as above. Then for fixed $\varepsilon > 0$ and $\forall m,n \ge N(\varepsilon)$,

$$\mathbb{P}(Y \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})) \ge \begin{cases} \Omega(1) & \text{if } \rho = \Omega(\sqrt{(m+n)\log(m+n)}) \\ \Omega(\exp(-(m+n)\log(m+n))) & \text{if } \rho = 0,\ \mu = \text{Counting or Leb.} \\ \exp(-mn^{1/2+\varepsilon} - nm^{1/2+\varepsilon}) & \text{if } \rho \ll 1 \text{ and } \mu \ll \text{Counting or Leb.} \end{cases}$$

Key Idea

- (Approximation by ML parametric model)

$$X \approx Y \sim \mu_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$$

That is,

$$\begin{aligned} &(1)\ \text{Conditional on } Y \in \mathcal{T}(\mathbf{r}, \mathbf{c}),\ Y \overset{d}{=} X \\ &(2)\ \mathbb{P}(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})) \gg (m+n)\log(m+n) \end{aligned}$$

Key Idea

- (Approximation by ML parametric model)

$$X \approx Y \sim \mu_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$$

That is,

$$\text{(1) Conditional on } Y \in \mathcal{T}(\mathbf{r}, \mathbf{c}),\ Y \overset{d}{=} X$$
$$\text{(2) } \mathbb{P}(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})) \gg (m+n)\log(m+n)$$

- (Concentration of ML parametric model around mean)

$$\mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}} = \big(\psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))\big)_{i,j}$$

Key Idea

- (Approximation by ML parametric model)

$$X \approx Y \sim \mu_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$$

That is,

$$(1)\ \text{Conditional on } Y \in \mathcal{T}(\mathbf{r},\mathbf{c}),\ Y \overset{d}{=} X$$
$$(2)\ \mathbb{P}(Y \in \mathcal{T}(\mathbf{r},\mathbf{c})) \gg (m+n)\log(m+n)$$

- (Concentration of ML parametric model around mean)

$$\mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}} = \big(\psi'(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j))\big)_{i,j}$$

$$\mathbb{P}\left(\|W_X - W_Z\|_\square \geq t \,\middle|\, X \in \mathcal{T}(\mathbf{r},\mathbf{c}) \right) \leq \mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r},\mathbf{c})\right)^{-1} \mathbb{P}(\|W_Y - W_Z\|_\square \geq t)$$

Open problems

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji

Open problems

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji
- Extend the theory for more general base measure for RM ensemble than i.i.d.?

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji
- Extend the theory for more general base measure for RM ensemble than i.i.d.?
- δ-tame marginal dist. in the Shrödinger bridge / Optimal transport theory?

$$\min_{\mathcal{H}\in\Pi(\nu_1,\nu_2)} D_{KL}(\mathcal{H}\,\|\,e^{-\gamma/\varepsilon}\mu_1\otimes\mu_2)$$

Phase transition (ν_1,ν_2) deviates away from (μ_1,μ_2)?

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji
- Extend the theory for more general base measure for RM ensemble than i.i.d.?
- δ-tame marginal dist. in the Shrödinger bridge / Optimal transport theory?

$$\min_{\mathcal{H}\in\Pi(\nu_1,\nu_2)} D_{KL}(\mathcal{H} \,\|\, e^{-\gamma/\varepsilon}\mu_1\otimes\mu_2)$$

  Phase transition (ν_1,ν_2) deviates away from (μ_1,μ_2)?

- Perturbation bound on Shrödinger bridge / Optimal transport map?

$$\|Z^{\mathbf{r},\mathbf{c}} - Z^{\mathbf{r}',\mathbf{c}'}\|_F^2 \leq C_\delta \|(\mathbf{r},\mathbf{c}) - (\mathbf{r}',\mathbf{c}')\|_1.$$

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji
- Extend the theory for more general base measure for RM ensemble than i.i.d.?
- δ-tame marginal dist. in the Shrödinger bridge / Optimal transport theory?

$$\min_{\mathcal{H}\in\Pi(\nu_1,\nu_2)} D_{KL}(\mathcal{H} \,\|\, e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2)$$

  Phase transition (ν_1, ν_2) deviates away from (μ_1, μ_2)?
- Perturbation bound on Shrödinger bridge / Optimal transport map?

$$\|Z^{\mathbf{r},\mathbf{c}} - Z^{\mathbf{r}',\mathbf{c}'}\|_F^2 \leq C_\delta \|(\mathbf{r},\mathbf{c}) - (\mathbf{r}',\mathbf{c}')\|_1.$$

- Limit theory for Shrödinger bridge / Optimal transport map?

Open problems

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji
- Extend the theory for more general base measure for RM ensemble than i.i.d.?
- δ-tame marginal dist. in the Shrödinger bridge / Optimal transport theory?

$$\min_{\mathcal{H}\in\Pi(\nu_1,\nu_2)} D_{KL}(\mathcal{H} \,\|\, e^{-\gamma/\varepsilon}\mu_1\otimes\mu_2)$$

  Phase transition (ν_1,ν_2) deviates away from (μ_1,μ_2)?
- Perturbation bound on Shrödinger bridge / Optimal transport map?

$$\|Z^{\mathbf{r},\mathbf{c}} - Z^{\mathbf{r}',\mathbf{c}'}\|_F^2 \le C_\delta \|(\mathbf{r},\mathbf{c}) - (\mathbf{r}',\mathbf{c}')\|_1.$$

- Limit theory for Shrödinger bridge / Optimal transport map?
- Large Deviations Principle? (For random graphs with given degree sequence, LDP is done by Dhara and Sen '22 [7])
  - Ongoing work with Sumit Mukherjee

Open problems

- Limit of empirical eigenvalue distribution for random matrices w/ given margins?
  - Ongoing work with Kyeongsik Nam and Hongchang Ji
- Extend the theory for more general base measure for RM ensemble than i.i.d.?
- δ-tame marginal dist. in the Shrödinger bridge / Optimal transport theory?

$$\min_{\mathcal{H}\in\Pi(\nu_1,\nu_2)} D_{KL}(\mathcal{H}\,\|\, e^{-\gamma/\varepsilon}\mu_1\otimes\mu_2)$$

  Phase transition (ν_1,ν_2) deviates away from (μ_1,μ_2)?
- Perturbation bound on Shrödinger bridge / Optimal transport map?

$$\|Z^{\mathbf{r},\mathbf{c}} - Z^{\mathbf{r}',\mathbf{c}'}\|_F^2 \le C_\delta \|(\mathbf{r},\mathbf{c}) - (\mathbf{r}',\mathbf{c}')\|_1.$$

- Limit theory for Shrödinger bridge / Optimal transport map?
- Large Deviations Principle? (For random graphs with given degree sequence, LDP is done by Dhara and Sen '22 [7])
  - Ongoing work with Sumit Mukherjee
- Condition on other statistics than row/column margin?
  - Ongoing work with William Powell (grad student)

Thank you very much!

[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. "On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries". In: *Advances in Mathematics* 224.1 (2010), pp. 316–339.

[3] Alexander Barvinok. "What does a random contingency table look like?" In: *Combinatorics, Probability and Computing* 19.4 (2010), pp. 517–539.

[4] Petter Brändén, Jonathan Leake, and Igor Pak. "Lower bounds for contingency tables via Lorentzian polynomials". In: *Israel Journal of Mathematics* 253.1 (2023), pp. 43–90.

[5] 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.

[6] Sourav Chatterjee, Persi Diaconis, and Allan Sly. "Random graphs with a given degree sequence". In: *The Annals of Applied Probability* 21.4 (2011), pp. 1400–1435.

[7] Souvik Dhara and Subhabrata Sen. "Large deviation for uniform graphs with given degrees". In: *Ann. Appl. Probab.* 32.3 (2022), pp. 2327–53.

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

[9] Samuel Dittmer, Hanbaek Lyu, and Igor Pak. "Phase transition in random contingency tables with non-uniform margins". In: *Transactions of the American Mathematical Society* 373.12 (2020), pp. 8313–8338.

[10] Robert Fortet. "Résolution d'un système d'équations de M. Schrödinger". In: *Journal de Mathématiques Pures et Appliquées* 19.1-4 (1940), pp. 83–105.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] Hanbaek Lyu and Sumit Muhkerjee. "Concentration and limit of large random matrices with given margins". In: *In preparation. (Prepreint available upon request)* (2024).

[17] Hanbaek Lyu and Igor Pak. "On the number of contingency tables and the independence heuristic". In: *Bulletin of the London Mathematical Society* 54.1 (2022), pp. 242–255.

[18] Michele Pavon, Giulio Trigila, and Esteban G Tabak. "The Data-Driven Schrödinger Bridge". In: *Communications on Pure and Applied Mathematics* 74.7 (2021), pp. 1545–1573.

[19] Cédric Villani. *Topics in optimal transportation*. Vol. 58. American Mathematical Soc., 2021.

Outline

How do we compute the typical table?

- Solve the dual (MLE) problem:

How do we compute the typical table?

- Solve the dual (MLE) problem:
  - $$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$

How do we compute the typical table?

- Solve the dual (MLE) problem:
  - $$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$
  - Strictly concave maximization in two variables $\boldsymbol{\alpha}, \boldsymbol{\beta}$

    $\rightarrow$ Alternating Maximization! (a.k.a. Nonlinear Gauss-Seidel or BCD)

$$\begin{cases} \boldsymbol{\alpha}_k \leftarrow \arg\max_{\boldsymbol{\alpha} \in \mathbb{R}^m} \; g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{k-1}) \\ \boldsymbol{\beta}_k \leftarrow \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^n} \; g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}). \end{cases}$$

How do we compute the typical table?

- Solve the dual (MLE) problem:
  - $$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$
  - Strictly concave maximization in two variables $\boldsymbol{\alpha}, \boldsymbol{\beta}$

    $\rightarrow$ Alternating Maximization! (a.k.a. Nonlinear Gauss-Seidel or BCD)

$$\begin{cases} \boldsymbol{\alpha}_k \leftarrow \arg\max_{\boldsymbol{\alpha} \in \mathbb{R}^m} \, g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{k-1}) \\ \boldsymbol{\beta}_k \leftarrow \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^n} \, g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}). \end{cases}$$

  - Finding critical points for the marginal problems, it reduces to

$$\begin{cases} \text{For } 1 \le i \le m,\ \boldsymbol{\alpha}_k(i) \leftarrow \text{unique } \alpha \text{ s.t. } \mathbf{r}(i) = \sum_{j=1}^{n} \psi'(\alpha + \boldsymbol{\beta}_{k-1}(j)), \\ \text{For } 1 \le j \le n,\ \boldsymbol{\beta}_k(j) \leftarrow \text{unique } \beta \text{ s.t. } \mathbf{c}(j) = \sum_{i=1}^{m} \psi'(\boldsymbol{\alpha}_k(i) + \beta). \end{cases}$$

How do we compute the typical table?

- Solve the dual (MLE) problem:
  - $$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$
  - Strictly concave maximization in two variables $\boldsymbol{\alpha}, \boldsymbol{\beta}$

    $\rightarrow$ Alternating Maximization! (a.k.a. Nonlinear Gauss-Seidel or BCD)

$$\begin{cases} \boldsymbol{\alpha}_k \leftarrow \arg\max_{\boldsymbol{\alpha}\in\mathbb{R}^m} \; g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{k-1}) \\ \boldsymbol{\beta}_k \leftarrow \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^n} \; g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}). \end{cases}$$

  - Finding critical points for the marginal problems, it reduces to

$$\begin{cases} \text{For } 1 \le i \le m,\ \boldsymbol{\alpha}_k(i) \leftarrow \text{unique } \alpha \text{ s.t. } \mathbf{r}(i) = \sum_{j=1}^n \psi'(\alpha + \boldsymbol{\beta}_{k-1}(j)), \\ \text{For } 1 \le j \le n,\ \boldsymbol{\beta}_k(j) \leftarrow \text{unique } \beta \text{ s.t. } \mathbf{c}(j) = \sum_{i=1}^m \psi'(\boldsymbol{\alpha}_k(i) + \beta). \end{cases}$$

  - For $\mu = \text{Poisson}(1)$ (Schrödinger bridge), $\psi'(x) = e^x$, so

$$\begin{cases} \text{For } 1 \le i \le m,\ \boldsymbol{\alpha}_k(i) \leftarrow \log(\mathbf{r}(i)) - \log\left(\sum_{j=1}^n \exp(\boldsymbol{\beta}_{k-1}(j))\right), \\ \text{For } 1 \le i \le n,\ \boldsymbol{\beta}_k(j) \leftarrow \log(\mathbf{c}(j)) - \log\left(\sum_{i=1}^m \exp(\boldsymbol{\alpha}_k(i))\right). \end{cases}$$

How do we compute the typical table?

- Solve the dual (MLE) problem:
  - $$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$
  - Strictly concave maximization in two variables $\boldsymbol{\alpha}, \boldsymbol{\beta}$

    $\rightarrow$ Alternating Maximization! (a.k.a. Nonlinear Gauss-Seidel or BCD)

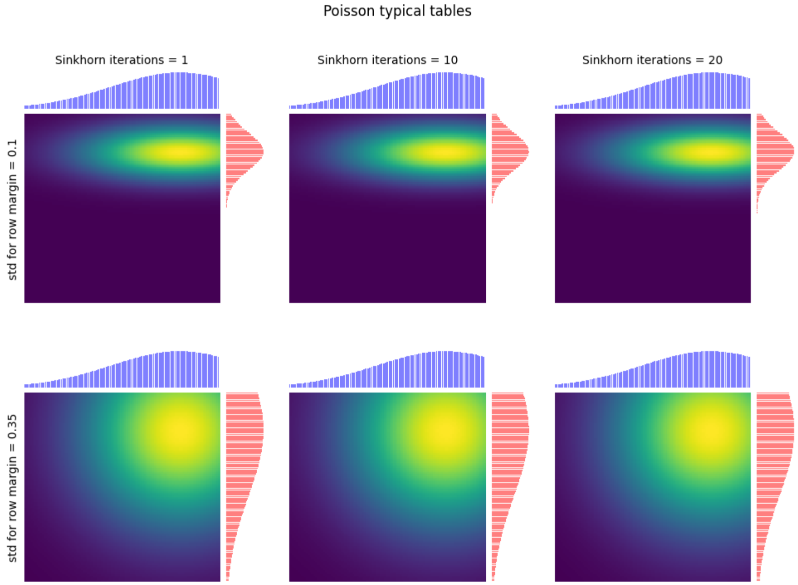
    $$\begin{cases} \boldsymbol{\alpha}_k \leftarrow \arg\max_{\boldsymbol{\alpha}\in\mathbb{R}^m} \; g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{k-1}) \\ \boldsymbol{\beta}_k \leftarrow \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^n} \; g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}). \end{cases}$$
  - Finding critical points for the marginal problems, it reduces to

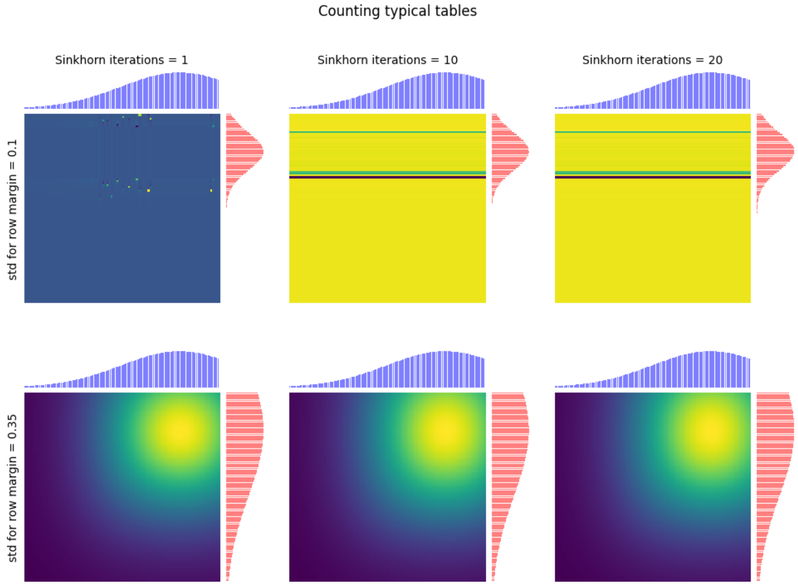
    $$\begin{cases} \text{For } 1 \le i \le m,\ \boldsymbol{\alpha}_k(i) \leftarrow \text{unique } \alpha \text{ s.t. } \mathbf{r}(i) = \sum_{j=1}^n \psi'(\alpha + \boldsymbol{\beta}_{k-1}(j)), \\ \text{For } 1 \le j \le n,\ \boldsymbol{\beta}_k(j) \leftarrow \text{unique } \beta \text{ s.t. } \mathbf{c}(j) = \sum_{i=1}^m \psi'(\boldsymbol{\alpha}_k(i) + \beta). \end{cases}$$
  - For $\mu = \text{Poisson}(1)$ (Schrödinger bridge), $\psi'(x) = e^x$, so

    $$\begin{cases} \text{For } 1 \le i \le m,\ \boldsymbol{\alpha}_k(i) \leftarrow \log(\mathbf{r}(i)) - \log\left(\sum_{j=1}^n \exp(\boldsymbol{\beta}_{k-1}(j))\right), \\ \text{For } 1 \le i \le n,\ \boldsymbol{\beta}_k(j) \leftarrow \log(\mathbf{c}(j)) - \log\left(\sum_{i=1}^m \exp(\boldsymbol{\alpha}_k(i))\right). \end{cases}$$

    Sinkhorn!!

Poisson and Counting Typical Tables




Counting typical tables
Sinkhorn iterations = 1
Sinkhorn iterations = 10
Sinkhorn iterations = 20
std for row margin = 0.1
std for row margin = 0.35

Theorem (Linear convergence of generalized Sinkhorn; **L**-Mukherjee '24+)

Fix μ arbitrary. Let $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$ = generalied Sinkhorn iterates. Fix an MLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for δ-tame $(\mathbf{r}, \mathbf{c})$ and denote $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$. Suppose ψ'' is monotonic and $\boldsymbol{\alpha}_0 = \mathbf{0}$ or μ admits arbitrary tilting ($\Theta = \mathbb{R}$). Then

$$\frac{\sigma_-(\varepsilon)^2}{2} \|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \leq \Delta_k \leq \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \geq 1.$$

Theorem (Linear convergence of generalized Sinkhorn; **L**-Mukherjee '24+)

Fix μ arbitrary. Let $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) =$ generalied Sinkhorn iterates. Fix an MLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for δ-tame $(\mathbf{r}, \mathbf{c})$ and denote $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$. Suppose ψ'' is monotonic and $\boldsymbol{\alpha}_0 = \mathbf{0}$ or μ admits arbitrary tilting ($\Theta = \mathbb{R}$). Then

$$\frac{\sigma_-(\varepsilon)^2}{2} \|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \leq \Delta_k \leq \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \geq 1.$$

- Key Challenges:

Theorem (Linear convergence of generalized Sinkhorn; **L**-Mukherjee '24+)

Fix μ arbitrary. Let $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) =$ generalied Sinkhorn iterates. Fix an MLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for δ-tame $(\mathbf{r}, \mathbf{c})$ and denote $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$. Suppose ψ'' is monotonic and $\boldsymbol{\alpha}_0 = \mathbf{0}$ or μ admits arbitrary tilting ($\Theta = \mathbb{R}$). Then

$$\frac{\sigma_-(\varepsilon)^2}{2}\|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \leq \Delta_k \leq \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \geq 1.$$

- Key Challenges:
  - The set of MLEs is unbounded: $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ MLE $\Longleftrightarrow$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ MLE $\forall \lambda \in \mathbb{R}$

Theorem (Linear convergence of generalized Sinkhorn; **L**-Mukherjee '24+)

Fix μ arbitrary. Let $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) =$ generalied Sinkhorn iterates. Fix an MLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for δ-tame $(\mathbf{r}, \mathbf{c})$ and denote $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$. Suppose ψ'' is monotonic and $\boldsymbol{\alpha}_0 = \mathbf{0}$ or μ admits arbitrary tilting ($\Theta = \mathbb{R}$). Then

$$\frac{\sigma_-(\varepsilon)^2}{2} \|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \leq \Delta_k \leq \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \geq 1.$$

- Key Challenges:
  - The set of MLEs is unbounded: $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ MLE $\Longleftrightarrow$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ MLE $\forall \lambda \in \mathbb{R}$
  - Need a priori bound on the Sinkhorn iterates

Theorem (Linear convergence of generalized Sinkhorn; **L**-Mukherjee '24+)

Fix μ *arbitrary. Let* $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) =$ *generalied Sinkhorn iterates. Fix an MLE* $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ *for* δ*-tame* $(\mathbf{r}, \mathbf{c})$ *and denote* $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$*. Suppose* ψ'' *is monotonic and* $\boldsymbol{\alpha}_0 = \mathbf{0}$ *or* μ *admits arbitrary tilting* ($\Theta = \mathbb{R}$)*. Then*

$$\frac{\sigma_-(\varepsilon)^2}{2} \|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \leq \Delta_k \leq \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \geq 1.$$

- Key Challenges:
  - The set of MLEs is unbounded: $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ MLE $\iff$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ MLE $\forall \lambda \in \mathbb{R}$
  - Need a priori bound on the Sinkhorn iterates

    $\Leftarrow$ For Schrödinger bridge ($\mu =$ Poisson(1)), use exact form of Sinkhorn updates

Theorem (Linear convergence of generalized Sinkhorn; **L**-Mukherjee '24+)

Fix μ arbitrary. Let $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$ = generalied Sinkhorn iterates. Fix an MLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for δ-tame $(\mathbf{r}, \mathbf{c})$ and denote $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$. Suppose ψ'' is monotonic and $\boldsymbol{\alpha}_0 = \mathbf{0}$ or μ admits arbitrary tilting ($\Theta = \mathbb{R}$). Then

$$\frac{\sigma_-(\varepsilon)^2}{2}\|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \leq \Delta_k \leq \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \geq 1.$$

- Key Challenges:
  - The set of MLEs is unbounded: $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ MLE $\Longleftrightarrow$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ MLE $\forall \lambda \in \mathbb{R}$
  - Need a priori bound on the Sinkhorn iterates

    $\Leftarrow$ For Schrödinger bridge (μ = Poisson(1)), use exact form of Sinkhorn updates
    For general μ, no exact form of Sinkhorn updates (implicit)

Theorem (Linear convergence of generalized Sinkhorn; L-Mukherjee '24+)

Fix μ arbitrary. Let $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) =$ generalied Sinkhorn iterates. Fix an MLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for δ-tame $(\mathbf{r}, \mathbf{c})$ and denote $\Delta_k := g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) - g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$. Suppose ψ'' is monotonic and $\boldsymbol{\alpha}_0 = \mathbf{0}$ or μ admits arbitrary tilting ($\Theta = \mathbb{R}$). Then

$$\frac{\sigma_-(\varepsilon)^2}{2} \|(\boldsymbol{\alpha}^* \oplus \boldsymbol{\beta}^*) - (\boldsymbol{\alpha}_k \oplus \boldsymbol{\beta}_k)\|_F^2 \le \Delta_k \le \left(1 - \frac{\sigma_-(\varepsilon)^4}{\sigma_+(\varepsilon)^4}\right)^{k-1} \Delta_1 \quad \text{for all } k \ge 1.$$

- Key Challenges:
  - The set of MLEs is unbounded: $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ MLE $\Longleftrightarrow$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ MLE $\forall \lambda \in \mathbb{R}$
  - Need a priori bound on the Sinkhorn iterates

    $\Leftarrow$ For Schrödinger bridge ($\mu =$ Poisson(1)), use exact form of Sinkhorn updates
    For general μ, no exact form of Sinkhorn updates (implicit)
- Solution: *We show the ℓ^∞-distance between the iterates and the set of MLEs does not expand*