Concentration and limit of large random matrices with given margins

Hanbaek Lyu

University of Wisconsin - Madison

Based on joint work with Sumit Mukherjee

June. 3, 2024

Outline

Random matrix with i.i.d. entries given margins

- *(Base model)* $\mu =$ probability measure on $\mathbb{R}$, 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 $\mathbb{R}$, 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 $\mathbf{x}$:

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

For each $\rho \geq 0$, let

$$\mathcal{T}_\rho(\mathbf{r},\mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m\times n} \,:\, \|r(\mathbf{x}) - \mathbf{r}\|_1 \leq \rho,\ \|c(\mathbf{x}) - \mathbf{c}\|_1 \leq \rho \right\} \quad (\mathcal{T}(\mathbf{r},\mathbf{c}) := \mathcal{T}_0(\mathbf{r},\mathbf{c}))$$

- *(Base model)* $\mu =$ probability measure on $\mathbb{R}$, 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 $\mathbf{x}$:

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

For each $\rho \geq 0$, let

$$\mathcal{T}_\rho(\mathbf{r}, \mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m \times n} \,:\, \|r(\mathbf{x}) - \mathbf{r}\|_1 \leq \rho, \ \|c(\mathbf{x}) - \mathbf{c}\|_1 \leq \rho \ \right\} \quad (\mathcal{T}(\mathbf{r}, \mathbf{c}) := \mathcal{T}_0(\mathbf{r}, \mathbf{c}))$$

- *(Main question)* For ρ 'small',

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

Random matrix with i.i.d. entries given margins

- *(Base model)* $\mu =$ probability measure on $\mathbb{R}$, and let

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

$X \sim \mu^{\otimes(m\times n)}$: $(m \times n)$ 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 $\mathbf{x}$:

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

For each $\rho \ge 0$, let

$$\mathcal{T}_\rho(\mathbf{r}, \mathbf{c}) := \left\{ \mathbf{x} \in \mathbb{R}^{m\times n} \,:\, \|r(\mathbf{x}) - \mathbf{r}\|_1 \le \rho,\ \|c(\mathbf{x}) - \mathbf{c}\|_1 \le \rho \right\} \quad (\mathcal{T}(\mathbf{r}, \mathbf{c}) := \mathcal{T}_0(\mathbf{r}, \mathbf{c}))$$

- *(Main question)* For ρ 'small',

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

- This question still makes sense if μ is not a probability measure (i.e., counting measure on $\mathbb{N}$), as long as the measure of $\mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$ under $\mu^{\otimes(m\times n)}$ is in $(0, \infty)$

- (*Main question*)

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

- High-level answer

- (*Main question*)

If we condition $X \sim \mu^{\otimes(m\times n)}$ on being in $\mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$, 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 $\mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$, 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 $(\mathbf{r}, \mathbf{c})$

Exponential tilting

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

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

Exponential tilting

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

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

  - Set of all allowed tilting parameters:

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

Exponential tilting

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

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

  - Set of all allowed tilting parameters:

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

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

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

Exponential tilting

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

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

  - Set of all allowed tilting parameters:

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

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

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

  - $\psi' : \Theta^\circ \to (A, B)$ is strictly increasing ($\triangleright$ `tilt2mean` function)

Exponential tilting

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

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

  - Set of all allowed tilting parameters:

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

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

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

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

Exponential tilting

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

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

  - Set of all allowed tilting parameters:

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

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

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

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

- 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 margins $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^m$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$. The **typical table** Z for margin $(\mathbf{r}, \mathbf{c})$ is

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

- Fix margins $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^m$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$. The **typical table** Z for margin $(\mathbf{r}, \mathbf{c})$ is

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

  - $g =$ strictly concave since $f'(x) = -\phi(x)$, $f''(x) = -\phi'(x) = \frac{-1}{\mathrm{Var}(\mu_{\phi(x)})} < 0$

Typical table

- Fix margins $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^m$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$. The **typical table** Z for margin $(\mathbf{r}, \mathbf{c})$ is

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

  - $g =$ strictly concave 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

- Fix margins $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^m$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$. The **typical table** Z for margin $(\mathbf{r}, \mathbf{c})$ is

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

  - $g =$ strictly concave 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.

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

- (*Informal result I*)

  $X \sim \mu^{\otimes(m\times n)}$ **conditioned on being in** $\mathcal{T}_{\rho}(\mathbf{r},\mathbf{c})$ **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 I*)

  $X \sim \mu^{\otimes(m\times n)}$ **conditioned on being in** $\mathcal{T}_\rho(\mathbf{r},\mathbf{c})$ **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 II*)

  $$\Big[\, X \sim \mu^{\otimes(m\times n)} \text{ conditioned on being in } \mathcal{T}_\rho(\mathbf{r},\mathbf{c}) \,\Big] \approx Y,$$

  where Y **has independent entries** $Y_{ij} \sim \mu_{\alpha(i)+\boldsymbol{\beta}(j)}$ **and** $\mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}}$

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 $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$

- 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 vector $\mathbf{x} \in \mathbb{R}^m$, denote $\bar{\mathbf{x}}$ = the corresponding **step function** $[0,1] \to \mathbb{R}$,

$$\bar{\mathbf{x}}(t) := \sum_{i=1}^{m} \mathbf{x}(i) \mathbf{1}\Big((i-1)/m < t \leq i/m \Big).$$

- 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 vector $\mathbf{x} \in \mathbb{R}^m$, denote $\bar{\mathbf{x}}$ = the corresponding **step function** $[0,1] \to \mathbb{R}$,

$$\bar{\mathbf{x}}(t) := \sum_{i=1}^{m} \mathbf{x}(i) \mathbf{1}\Big((i-1)/m < t \le i/m\Big).$$

- 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 **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 vector $\mathbf{x} \in \mathbb{R}^m$, denote $\bar{\mathbf{x}}$ = the corresponding **step function** $[0,1] \to \mathbb{R}$,

$$\bar{\mathbf{x}}(t) := \sum_{i=1}^{m} \mathbf{x}(i) \mathbf{1}\Big((i-1)/m < t \leq i/m\Big).$$

- 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

$$\lim_{m,n\to\infty} \|\mathbf{r} - n^{-1}\bar{\mathbf{r}}_m\|_1 + \|\mathbf{c} - m^{-1}\bar{\mathbf{c}}_n\|_1 = 0.$$

- 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 vector $\mathbf{x} \in \mathbb{R}^m$, denote $\bar{\mathbf{x}}$ = the corresponding **step function** $[0,1] \to \mathbb{R}$,

$$\bar{\mathbf{x}}(t) := \sum_{i=1}^{m} \mathbf{x}(i) \mathbf{1}\Big((i-1)/m < t \leq i/m\Big).$$

- 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

$$\lim_{m,n\to\infty} \|\mathbf{r} - n^{-1}\bar{\mathbf{r}}_m\|_1 + \|\mathbf{c} - m^{-1}\bar{\mathbf{c}}_n\|_1 = 0.$$

- (*Informal result III*)

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

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

1. Connection to Entropy minimization and Schrödinger bridge

Plan of the talk

1. Connection to Entropy minimization and Schrödinger bridge
2. Connection to Random graphs with given degree sequence

Plan of the talk

1. Connection to Entropy minimization and Schrödinger bridge
2. Connection to Random graphs with given degree sequence
3. Connection to Contingency tables and Phase transition

Plan of the talk

1. Connection to Entropy minimization and Schrödinger bridge
2. Connection to Random graphs with given degree sequence
3. Connection to Contingency tables and Phase transition
4. Key ideas and Formal statement of results

Plan of the talk

1. Connection to Entropy minimization and Schrödinger bridge
2. Connection to Random graphs with given degree sequence
3. Connection to Contingency tables and Phase transition
4. Key ideas and Formal statement of results
5. Open problems

- (*Base model*) Reference probability measure $\mathcal{R}$ on $\mathbb{R}^{m\times n}$ for $m \times n$ random matrices

Relative entropy minimization given expected margins

- (*Base model*) Reference probability measure $\mathcal{R}$ on $\mathbb{R}^{m\times n}$ for $m\times n$ random matrices
- (*Relative entropy*) For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of $\mathcal{H}$ from $\mathcal{R}$* is

$$D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) := \begin{cases} \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\,\mathcal{R}(d\mathbf{x}) & \textit{if } \mathcal{H}\ll\mathcal{R} \\ \infty & \text{otherwise.} \end{cases}$$

- *(Base model)* Reference probability measure $\mathcal{R}$ on $\mathbb{R}^{m\times n}$ for $m \times n$ random matrices
- *(Relative entropy)* For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of* $\mathcal{H}$ *from* $\mathcal{R}$ is

$$D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) := \begin{cases} \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\, \mathcal{R}(d\mathbf{x}) & \textit{if } \mathcal{H} \ll \mathcal{R} \\ \infty & \text{otherwise.} \end{cases}$$

- Find the least-action distribution $\mathcal{H}$ that minimizes the relative entropy from $\mathcal{R}$ constrained on the expected margins being $(\mathbf{r}, \mathbf{c})$:

$$\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X), c(X))] = (\mathbf{r}, \mathbf{c})$$

where $\mathcal{P}^{m\times n}$ = the set of all probability measures on $\mathbb{R}^{m\times n}$

- (*Base model*) Reference probability measure $\mathcal{R}$ on $\mathbb{R}^{m\times n}$ for $m \times n$ random matrices
- (*Relative entropy*) For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of* $\mathcal{H}$ *from* $\mathcal{R}$ is

$$D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) := \begin{cases} \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\, \mathcal{R}(d\mathbf{x}) & \textit{if } \mathcal{H} \ll \mathcal{R} \\ \infty & \text{otherwise.} \end{cases}$$

- Find the least-action distribution $\mathcal{H}$ that minimizes the relative entropy from $\mathcal{R}$ constrained on the expected margins being $(\mathbf{r}, \mathbf{c})$:

$$\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X), c(X))] = (\mathbf{r}, \mathbf{c})$$

where $\mathcal{P}^{m\times n}$ = the set of all probability measures on $\mathbb{R}^{m\times n}$

  - Analogous to maximum entropy probability distributions given moments

- *(Base model)* Reference probability measure $\mathcal{R}$ on $\mathbb{R}^{m\times n}$ for $m \times n$ random matrices
- *(Relative entropy)* For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of $\mathcal{H}$ from $\mathcal{R}$* is

$$D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) := \begin{cases} \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\, \mathcal{R}(d\mathbf{x}) & \textit{if } \mathcal{H} \ll \mathcal{R} \\ \infty & \text{otherwise.} \end{cases}$$

- Find the least-action distribution $\mathcal{H}$ that minimizes the relative entropy from $\mathcal{R}$ constrained on the expected margins being $(\mathbf{r}, \mathbf{c})$:

$$\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X), c(X))] = (\mathbf{r}, \mathbf{c})$$

where $\mathcal{P}^{m\times n}$ = the set of all probability measures on $\mathbb{R}^{m\times n}$

  - Analogous to maximum entropy probability distributions given moments
  - e.g.,

$$\begin{aligned} \mathrm{Exp}(\lambda) &= \operatorname*{arg\,min}_{\mathcal{H}\in\mathcal{P}(\mathbb{R})} D_{KL}(\mathcal{H}\,|\,\mathrm{Leb}(\mathbb{R}_{\geq 0})) \quad \text{s.t.}\, \mathbb{E}[\mathcal{H}] = \lambda \\ &= \operatorname*{arg\,max}_{h} - \int_{\mathbb{R}} h(x)\log h(x)\, dx \quad \text{s.t.} \int_{\mathbb{R}} x\, h(x)\, dx = \lambda \end{aligned}$$

Relative entropy minimization given expected margins

- $(*)$ $$\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X), c(X))] = (\mathbf{r}, \mathbf{c}).$$

Relative entropy minimization given expected margins

- (*) $$\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X),c(X))] = (\mathbf{r},\mathbf{c}).$$
- Specializing to $\mathcal{R}=\mu^{\otimes(m\times n)}$, the optimal measure $\mathcal{H}_{\text{opt}}$ is the product of some entrywise exponential tilting $\mu_{\theta_{ij}}$: $\mathcal{H}_{\text{opt}}=\bigotimes_{i,j}\mu_{\theta_{i,j}}$.

- $(*)$ $\quad \min_{\mathcal{H} \in \mathcal{P}^{m\times n}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad$ subject to $\quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X), c(X))] = (\mathbf{r}, \mathbf{c})$.
- Specializing to $\mathcal{R} = \mu^{\otimes(m\times n)}$, the optimal measure $\mathcal{H}_{\text{opt}}$ is the product of some entrywise exponential tilting $\mu_{\theta_{ij}}$: $\mathcal{H}_{\text{opt}} = \bigotimes_{i,j} \mu_{\theta_{i,j}}$.
  - Convert to an OPT problem in $h = \frac{d\mathcal{H}}{d\mathcal{R}}$

- $(*)$ $\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R})$ subject to $\mathbb{E}_{X\sim\mathcal{H}}[(r(X),c(X))] = (\mathbf{r},\mathbf{c})$.
- Specializing to $\mathcal{R}=\mu^{\otimes(m\times n)}$, the optimal measure $\mathcal{H}_{\text{opt}}$ is the product of some entrywise exponential tilting $\mu_{\theta_{ij}}$: $\mathcal{H}_{\text{opt}} = \bigotimes_{i,j}\mu_{\theta_{i,j}}$.

  - Convert to an OPT problem in $h = \frac{d\mathcal{H}}{d\mathcal{R}}$
  - Then the corresponding Lagrangian is

$$\mathcal{L}(h) = \int_{\mathbb{R}^{m\times n}} h(\mathbf{x})\log\left(h(\mathbf{x})\right)\, R(d\mathbf{x}) + \lambda\left(\int_{\mathbb{R}^{m\times n}} h(\mathbf{x})\, d\mathcal{R}(\mathbf{x}) - 1\right)$$

$$+\sum_{i=1}^{m}\alpha_i\left(\sum_{j=1}^{n}\int_{\mathbb{R}^{m\times n}} x_{ij}h(\mathbf{x})\, d\mathcal{R}(\mathbf{x}) - \mathbf{r}(i)\right) + \sum_{j=1}^{n}\beta_j\left(\sum_{i=1}^{m}\int_{\mathbb{R}^{m\times n}} x_{ij}h(\mathbf{x})\, d\mathcal{R}(\mathbf{x}) - \mathbf{c}(j)\right)$$

Relative entropy minimization given expected margins

- $(*)\qquad \min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R})$ subject to $\mathbb{E}_{X\sim\mathcal{H}}[(r(X),c(X))]=(\mathbf{r},\mathbf{c})$.
- Specializing to $\mathcal{R}=\mu^{\otimes(m\times n)}$, the optimal measure $\mathcal{H}_{\text{opt}}$ is the product of some entrywise exponential tilting $\mu_{\theta_{ij}}$: $\mathcal{H}_{\text{opt}}=\bigotimes_{i,j}\mu_{\theta_{i,j}}$.

  - Convert to an OPT problem in $h=\frac{d\mathcal{H}}{d\mathcal{R}}$
  - Then the corresponding Lagrangian is

$$\mathcal{L}(h)=\int_{\mathbb{R}^{m\times n}} h(\mathbf{x})\log\left(h(\mathbf{x})\right)\,R(d\mathbf{x})+\lambda\left(\int_{\mathbb{R}^{m\times n}} h(\mathbf{x})\,d\mathcal{R}(\mathbf{x})-1\right)$$

$$+\sum_{i=1}^{m}\alpha_i\left(\sum_{j=1}^{n}\int_{\mathbb{R}^{m\times n}} x_{ij}h(\mathbf{x})\,d\mathcal{R}(\mathbf{x})-\mathbf{r}(i)\right)+\sum_{j=1}^{n}\beta_j\left(\sum_{i=1}^{m}\int_{\mathbb{R}^{m\times n}} x_{ij}h(\mathbf{x})\,d\mathcal{R}(\mathbf{x})-\mathbf{c}(j)\right)$$

  - Suppose the sol. of $(*)$ is attained in the interior. Then $\frac{\delta\mathcal{L}}{\delta h}=0$, so

$$\log h_{\text{opt}}(\mathbf{x})+1+\lambda+x_{ij}\alpha_i+x_{ij}\beta_j=0\quad\text{for all } i,j\quad \mathcal{R}\text{-a.s.}\ ,$$

$$\log h_{\text{opt}}(\mathbf{x})\propto\sum_{ij}(\alpha_i+\beta_{j,})x_{ij}\qquad \mathcal{R}\text{-a.s.}$$

Relative entropy minimization given expected margins

- Specializing to $\mathcal{R} = \mu^{\otimes(m\times n)}$, the optimal measure $\mathcal{H}$ is a product of entrywise exponential tilting $\mu_{\theta_{ij}}$:

$$\mathcal{H}_{\mathrm{opt}} = \bigotimes_{i,j} \mu_{\theta_{i,j}}$$

- Specializing to $\mathcal{R} = \mu^{\otimes(m \times n)}$, the optimal measure $\mathcal{H}$ is a product of entrywise exponential tilting $\mu_{\theta_{ij}}$:

$$\mathcal{H}_{\text{opt}} = \bigotimes_{i,j} \mu_{\theta_{i,j}}$$

- Parameterize entrywise exponential tilting by the **mean matrix** $Z = (z_{ij})$:

$$\mathcal{H}_{\text{opt}} = \bigotimes_{i,j} \mu_{\phi(z_{ij})}$$

Relative entropy minimization given expected margins

- Specializing to $\mathcal{R} = \mu^{\otimes(m\times n)}$, the optimal measure $\mathcal{H}$ is a product of entrywise exponential tilting $\mu_{\theta_{ij}}$:

$$\mathcal{H}_{\mathrm{opt}} = \bigotimes_{i,j} \mu_{\theta_{i,j}}$$

- Parameterize entrywise exponential tilting by the **mean matrix** $Z = (z_{ij})$:

$$\mathcal{H}_{\mathrm{opt}} = \bigotimes_{i,j} \mu_{\phi(z_{ij})}$$

- Therefore the relative entropy minimization problem

$$\min_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X),c(X))] = (\mathbf{r},\mathbf{c}).$$

reduces to

$$\min_{Z\in\mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{D_{KL}\left(\bigotimes_{i,j}\mu_{\phi(z_{ij})} \,\middle\|\, \bigotimes_{i,j}\mu\right)}_{=\sum_{i,j} D(\mu_{\phi(z_{ij})}\,\|\,\mu)} \qquad (\triangleright \textit{ Typical table problem!})$$

- Specialze to $(m \times n) = (1 \times 2)$, 2-dim random vector $X = (X_1, X_2) \sim \mathcal{R}$. Constrain the marginal distributions, not the expected (row/column) margin:

$$\min_{\mathcal{H} \in \mathcal{P}^{1\times 2}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{subject to} \quad X_1 =_d \mu_1,\, X_2 =_d \mu_2 \text{ where } (X_1, X_2) \sim \mathcal{H}.$$

- Specialze to $(m \times n) = (1 \times 2)$, 2-dim random vector $X = (X_1, X_2) \sim \mathcal{R}$. Constrain the marginal distributions, not the expected (row/column) margin:

$$\min_{\mathcal{H} \in \mathcal{P}^{1\times 2}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{subject to} \quad X_1 =_d \mu_1,\, X_2 =_d \mu_2 \text{ where } (X_1, X_2) \sim \mathcal{H}.$$

- In a more familar form,

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

The optimal $\mathcal{H}$ from above is the **static Schrodinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$

- Specialze to $(m \times n) = (1 \times 2)$, 2-dim random vector $X = (X_1, X_2) \sim \mathcal{R}$. Constrain the marginal distributions, not the expected (row/column) margin:

$$\min_{\mathcal{H} \in \mathcal{P}^{1\times 2}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{subject to} \quad X_1 =_d \mu_1,\ X_2 =_d \mu_2 \text{ where } (X_1, X_2) \sim \mathcal{H}.$$

- In a more familar form,

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

  The optimal $\mathcal{H}$ from above is the **static Schrodinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$

- Specialize to $\mathcal{R} = \mu_1 \otimes \mu_2$. Then $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Shrödinger potentials** [14], characterizing the Shrödinger bridge as

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

  (Lagrange multiplier for the marginal distribution constraint $X_i =_d \mu_i$)

- Specialze to $(m \times n) = (1 \times 2)$, 2-dim random vector $X = (X_1, X_2) \sim \mathcal{R}$. Constrain the marginal distributions, not the expected (row/column) margin:

$$\min_{\mathcal{H} \in \mathcal{P}^{1\times 2}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{subject to} \quad X_1 =_d \mu_1,\ X_2 =_d \mu_2 \text{ where } (X_1, X_2) \sim \mathcal{H}.$$

- In a more familar form,

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

  The optimal $\mathcal{H}$ from above is the **static Schrodinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$

- Specialize to $\mathcal{R} = \mu_1 \otimes \mu_2$. Then $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Shrödinger potentials** [14], characterizing the Shrödinger bridge as

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

  (Lagrange multiplier for the marginal distribution constraint $X_i =_d \mu_i$)

  - Typical table $\leftarrow$ joint exponential tilting (from the first-moment condition),

- Specialze to $(m \times n) = (1 \times 2)$, 2-dim random vector $X = (X_1, X_2) \sim \mathcal{R}$. Constrain the marginal distributions, not the expected (row/column) margin:

$$\min_{\mathcal{H} \in \mathcal{P}^{1\times 2}} D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{subject to} \quad X_1 =_d \mu_1,\ X_2 =_d \mu_2 \text{ where } (X_1, X_2) \sim \mathcal{H}.$$

- In a more familar form,

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

The optimal $\mathcal{H}$ from above is the **static Schrodinger bridge** between μ_1 and μ_2 w.r.t. $\mathcal{R}$

- Specialize to $\mathcal{R} = \mu_1 \otimes \mu_2$. Then $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Shrödinger potentials** [14], characterizing the Shrödinger bridge as

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

(Lagrange multiplier for the marginal distribution constraint $X_i =_d \mu_i$)

- Typical table $\leftarrow$ joint exponential tilting (from the first-moment condition),
- Schrödinger bridge $\leftarrow$ reweighting (from entrywise distributional constraint).

Outline

Introduction

Entropy minimization and Shrödinger bridge

Random graphs with given degree sequences

Contingency tables and Phase transition

Statement of results

Open problems

- $\mathbf{d} = (d_1, d_2, \ldots, d_n)$: **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?

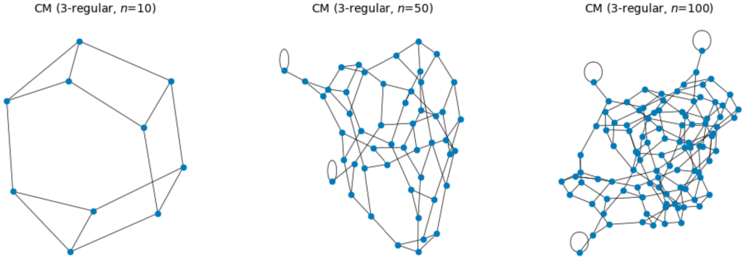


Figure: Random 3-regular graphs generated by the configuration model (allowing loops)

- $(\mathbf{d}^n)_{n\geq 1}$: 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

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

    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

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

  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)

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

- 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:

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

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)

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

- 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)

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

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

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

- **Sketch of proof**:

- **Sketch of proof**:
  - Find MLE β^n for the $\boldsymbol{\beta}$-model to the target degree sequence $\mathbf{d}^n$

Key idea in CDS '11 [6]

- **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

- **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:

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

- **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)

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

- **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 the target deg. seq. $\mathbf{d}^n$ with prob. $\geq \exp(-o(n^{2/3+\varepsilon}))$

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

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

Outline

- 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

$X = (X_{ij})$

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

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

- 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

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

- **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]):

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

Uniform and smooth margins

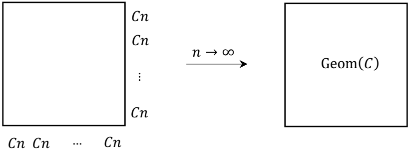
- **Uniform margins**: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.
  - Sharp volume estimate (Canfield and MacKay '10 [4]):

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

  - 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 and smooth margins

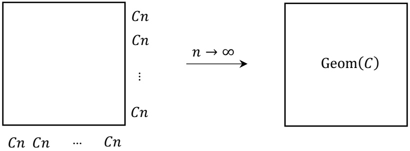
- **Uniform margins**: $\mathbf{a} = \mathbf{b} = (\lfloor Cn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$.
  - Sharp volume estimate (Canfield and MacKay '10 [4]):

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

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

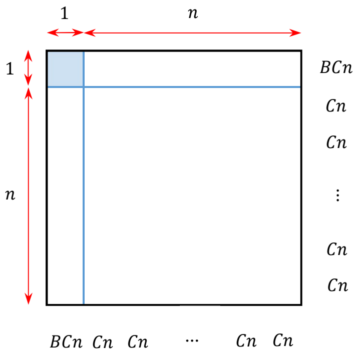
$$d_{TV}(X_{ij}, \mathrm{Geom}(C)) \to 0 \quad \text{as } n \to \infty$$

  Asymptotically independent entries

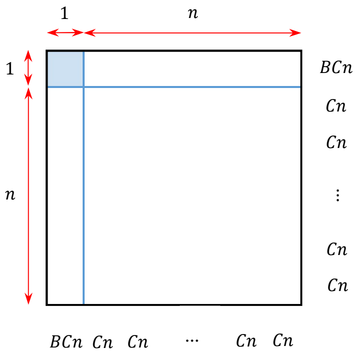


  - Empirical distribution of eigenvalues $\Rightarrow$ circular law (Nguyen '14 [12])

- What about non-uniform margins?

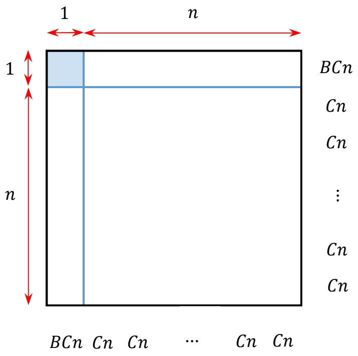


- What about non-uniform margins?



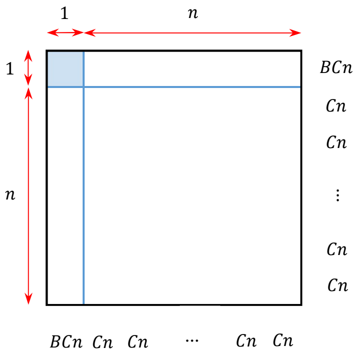
- Let $\mathbf{r} = \mathbf{c} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$. Let $X = (X_{ij})$ be the uniform contingency table with this margin.

- What about non-uniform margins?



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

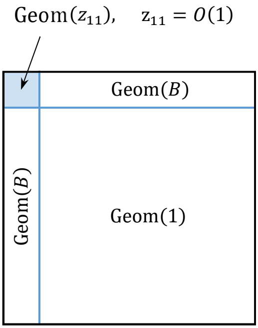
- What about non-uniform margins?



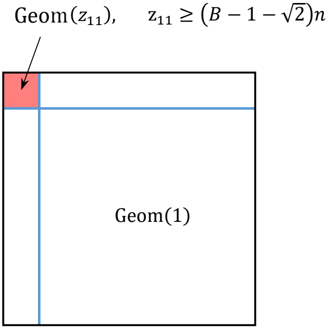
- Let $\mathbf{r} = \mathbf{c} = (\lfloor BCn \rfloor, \lfloor Cn \rfloor, \cdots, \lfloor Cn \rfloor) \in \mathbb{N}^n$. Let $X = (X_{ij})$ be the uniform contingency table with this margin.
- Do we still have convergence to geometric entries for all $B, C \geq 1$?
- If so, what are the means of the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

Barvinok's conjecture

- Based on a **typical table** computation, Barvinok conjectured in 2010 that each entry in X is asymptotically distributed as a geometric variable.

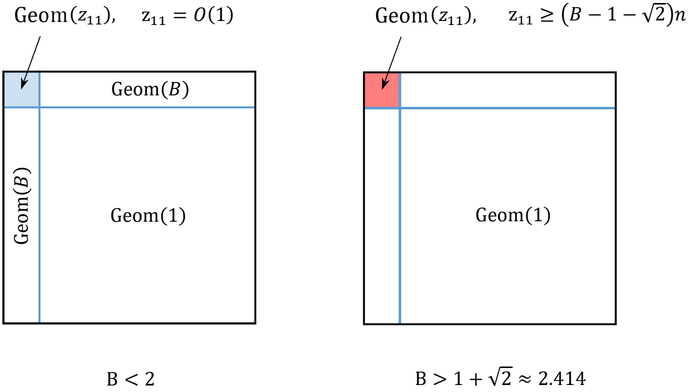


$B < 2$



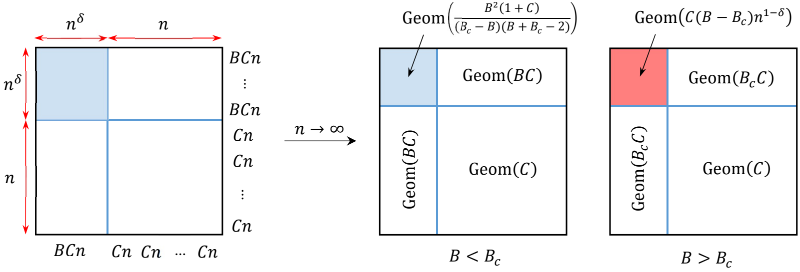
$B > 1 + \sqrt{2} \approx 2.414$

- Based on a **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 $B > 1 + \sqrt{2}$.



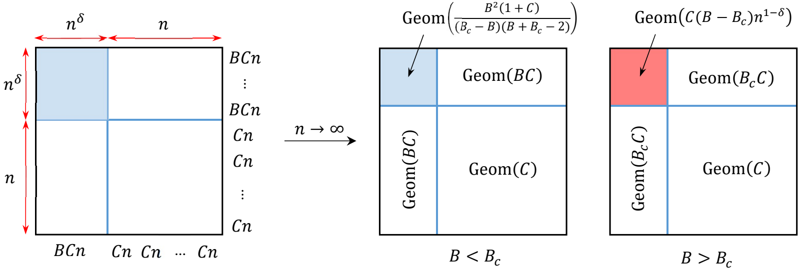
Theorem (Dittmer, L., and Pak '20)

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:



- **Where is this sharp phase transition coming from?**

- (Barvinok '10 [2]) For a $m \times n$ margin $(\mathbf{r}, \mathbf{c})$, the corresponding typical table is

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

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

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

- (Barvinok '10 [2]) For a $m \times n$ margin $(\mathbf{r}, \mathbf{c})$, the corresponding typical table is

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

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

$$= \underset{X \in \mathcal{T}(\mathbf{r},\mathbf{c})}{\arg\min} \sum_{i,j} D\left(\mu_{\phi(x_{ij})} \,\|\, \mu = \mathrm{Counting}(\mathbb{N})\right)$$

- Barvinok's insight:

Uniformly random CT with margin $(\mathbf{r}, \mathbf{c}) \approx Z^{\mathbf{r},\mathbf{c}}$

- (Barvinok '10 [2]) For a $m \times n$ margin $(\mathbf{r}, \mathbf{c})$, the corresponding typical table is

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

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

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

- Barvinok's insight:

Uniformly random CT with margin $(\mathbf{r}, \mathbf{c}) \approx Z^{\mathbf{r},\mathbf{c}}$

- (Barvinok '09 [1], '10 [2]) $Y = (n \times n)$ random matrix of independent entries $Y_{ij} \sim \mathrm{Geom}(z_{ij}) = \mu_{\phi(z_{ij})}$. Then for each CT X with margin $(\mathbf{r}, \mathbf{c})$,

$$\mathbb{P}(Y = X \mid Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})) \equiv \exp(-g(Z))$$

- (Barvinok '10 [2]) For a $m \times n$ margin $(\mathbf{r}, \mathbf{c})$, the corresponding typical table is

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

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

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

- Barvinok's insight:

Uniformly random CT with margin $(\mathbf{r}, \mathbf{c}) \approx Z^{\mathbf{r},\mathbf{c}}$

  - (Barvinok '09 [1], '10 [2]) $Y = (n \times n)$ random matrix of independent entries $Y_{ij} \sim \text{Geom}(z_{ij}) = \mu_{\phi(z_{ij})}$. Then for each CT X with margin $(\mathbf{r}, \mathbf{c})$,

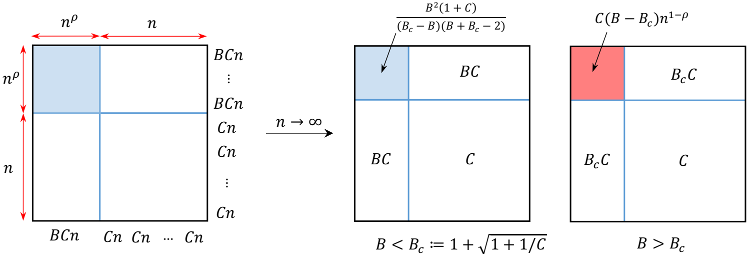
$$\mathbb{P}(Y = X \mid Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})) \equiv \exp(-g(Z))$$

  - Brändén, Leake, and Pak '23 [3] generalized this result to CTs with possibly bounded integer values (Using Lorenzian polynomials)

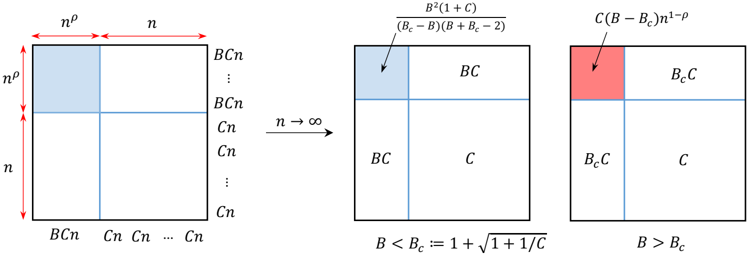
Sharp phase transition in typical tables

- Typical tables can change drastically by a small change in the margin!

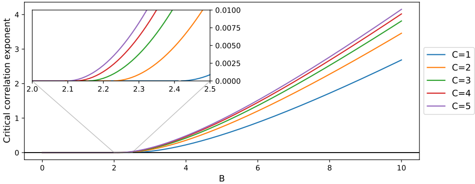
- Typical tables can change drastically by a small change in the margin!
  - For $0 \leq \delta < 1$, DLP [8] show that $Z^{\mathbf{r},\mathbf{c}}$ undergoes a **sharp phase transition** at $B_c = 1 + \sqrt{1 + C^{-1}}$:



- Typical tables can change drastically by a small change in the margin!
  - For $0 \le \delta < 1$, DLP [8] 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 [10]



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

Outline

What have we learned?

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

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

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

What have we learned?

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

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

- A natural statistical model for margin-conditioned RM $=$ $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model

What have we learned?

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

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

- A natural statistical model for margin-conditioned RM $=$ $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij})$, independent entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$

What have we learned?

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

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

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

What have we learned?

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

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

- A natural statistical model for margin-conditioned RM $=$ $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij})$, 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 [9])

What have we learned?

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

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

- A natural statistical model for margin-conditioned RM $=$ $(\boldsymbol{\alpha},\boldsymbol{\beta})$-model
  - $Y = (Y_{ij})$, 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 [9])
  - $\operatorname*{arg\,min}_{Z} \underline{\texttt{Relative Entropy}}(Z) = \operatorname*{arg\,max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \underline{\texttt{Likelihood Of Margin}}(\boldsymbol{\alpha},\boldsymbol{\beta})$

What have we learned?

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

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

- A natural statistical model for margin-conditioned RM = $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model
  - $Y = (Y_{ij})$, 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 [9])
  - $$\operatorname*{arg\,min}_{Z} \underline{\texttt{Relative Entropy}}(Z) = \operatorname*{arg\,max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \underline{\texttt{Likelihood Of Margin}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$
  - For $\psi' = \texttt{tilt2mean}$ function,

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

Tame margins

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

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

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

(i.e., corresponding MLE is δ-non-degenearte)

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

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

(i.e., corresponding MLE is δ-non-degenearte)

- Tanemess for μ with compact support $\iff$ Boundedness of MLE

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

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

(i.e., corresponding MLE is δ-non-degenearte)

  - Tanemess for μ with compact support $\iff$ Boundedness of MLE
  - Non-tanemess for μ with unbounded support $\iff$ Diverging typical table

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

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

(i.e., corresponding MLE is δ-non-degenearte)

  - Tanemess for μ with compact support $\iff$ Boundedness of MLE
  - Non-tanemess for μ with unbounded support $\iff$ Diverging typical table
- (**Sharp conditions for tameness**; this work [9])

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

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

  (i.e., corresponding MLE is δ-non-degenearte)
  - Tanemess for μ with compact support $\iff$ Boundedness of MLE
  - Non-tanemess for μ with unbounded support $\iff$ Diverging typical table
- (**Sharp conditions for tameness**; this work [9])
  - Suppose $sn \le \mathbf{r}_- \le \mathbf{r}_+ \le tn$ and $sm \le \mathbf{c}_- \le \mathbf{c}_+ \le tm$. Then δ-tame if $\frac{t}{s} \ll 1$

Tame margins

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

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

  (i.e., corresponding MLE is δ-non-degenearte)
  - Tanemess for μ with compact support $\Longleftrightarrow$ Boundedness of MLE
  - Non-tanemess for μ with unbounded support $\Longleftrightarrow$ Diverging typical table
- (**Sharp conditions for tameness**; this work [9])
  - Suppose $sn \le \mathbf{r}_- \le \mathbf{r}_+ \le tn$ and $sm \le \mathbf{c}_- \le \mathbf{c}_+ \le tm$. Then δ-tame if $\frac{t}{s} \ll 1$
  - In particular, for $\mu = \text{Counting}(\mathbb{N})$, then

$$\begin{aligned} \delta\text{-tame} \quad &\Longleftarrow \quad t/s < \rho_c := 1+\sqrt{1+s^{-1}} \\ \text{Diverging typical tables} \quad &\Longleftarrow \quad t/s > \rho_c \end{aligned}$$

  (Generalizes the subcritical behavior of Barvinok margin) ($\rho_c = B_c$ for $s = C$)

Theorem (L., and Muhkerjee '24+)

$(\mathbf{r},\mathbf{c}) = (m \times n)$ *δ-tame margin,* $Z = Z^{\mathbf{r},\mathbf{c}}$ *Let* $X = (X_{ij}) \sim \mu^{\otimes(m\times n)}$*. Then for some constants* $C_i = C_i(\delta,\mu) > 0$ *for* $i = 1, 2$*,*

(i) $\exists$ *MLE* $\boldsymbol{\alpha} \in \mathbb{R}^m$*,* $\boldsymbol{\beta} \in \mathbb{R}^n$ *such that* $Z_{ij} = \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))$ *for all* i, j*.*

Theorem (L., and Muhkerjee '24+)

$(\mathbf{r}, \mathbf{c}) = (m \times n)$ *δ-tame margin,* $Z = Z^{\mathbf{r},\mathbf{c}}$ *Let* $X = (X_{ij}) \sim \mu^{\otimes(m\times n)}$. *Then for some constants* $C_i = C_i(\delta, \mu) > 0$ *for* $i = 1, 2$,

(i) $\exists$ *MLE* $\boldsymbol{\alpha} \in \mathbb{R}^m$, $\boldsymbol{\beta} \in \mathbb{R}^n$ *such that* $Z_{ij} = \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))$ *for all* i, j.

(ii) *Let* $Y = (Y_{ij})$ *be a* $m \times n$ *RM of indep. entries* $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$. *Then* $\mathbb{E}[Y] = Z$ *and for each* $t, \varepsilon \geq 0$,

$$\mathbb{P}\left(\|W_X - W_Z\|_\square \geq t \,\middle|\, X \in \mathcal{T}_{\varepsilon mn}(\mathbf{r}, \mathbf{c}) \right)$$
$$\leq \mathbb{P}\left(Y \in \mathcal{T}_{\varepsilon mn}(\mathbf{r}, \mathbf{c})\right)^{-1} \exp\left(C_1 \varepsilon mn + (m+n+1)\log 2 - \frac{t^2 mn}{2C_1} \right).$$

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

Theorem (L., and Muhkerjee '24+)

$(\mathbf{r}, \mathbf{c}) = (m \times n)$ δ-*tame margin*, $Z = Z^{\mathbf{r},\mathbf{c}}$ *Let* $X = (X_{ij}) \sim \mu^{\otimes (m\times n)}$. *Then for some constants* $C_i = C_i(\delta, \mu) > 0$ *for* $i = 1, 2$,

(i) $\exists$ *MLE* $\boldsymbol{\alpha} \in \mathbb{R}^m$, $\boldsymbol{\beta} \in \mathbb{R}^n$ *such that* $Z_{ij} = \psi'(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j))$ *for all* i, j.

(ii) *Let* $Y = (Y_{ij})$ *be a* $m \times n$ *RM of indep. entries* $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$. *Then* $\mathbb{E}[Y] = Z$ *and for each* $t, \varepsilon \geq 0$,

$$\mathbb{P}\left(\|W_X - W_Z\|_\square \geq t \,\middle|\, X \in \mathcal{T}_{\varepsilon mn}(\mathbf{r}, \mathbf{c}) \right) \leq \mathbb{P}\left(Y \in \mathcal{T}_{\varepsilon mn}(\mathbf{r}, \mathbf{c})\right)^{-1} \exp\left(C_1 \varepsilon mn + (m+n+1)\log 2 - \frac{t^2 mn}{2C_1} \right).$$

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

(iii) *Denote* $s(m,n) := mn\sqrt{m^{-1} + n^{-1}}$. *Then*

$$\mathbb{P}\left(\|W_X - W_Z\|_\square \geq C_1 \left(m^{-1} + n^{-1}\right)^{1/4} \,\middle|\, X \in \mathcal{T}_{C_1 s(m,n)}(\mathbf{r}, \mathbf{c}) \right) \leq \exp\left(-C_2 s(m,n)\right).$$

Theorem (L., and Muhkerjee '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 *as* $m, n \to \infty$.

(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 $(\mathbf{r}, \mathbf{c})$.

Theorem (L., and Muhkerjee '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 as $m, n \to \infty$.

(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) *For $C = C(\mu, \delta) > 0$,*

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

In particular, $\|W^{\mathbf{r},\mathbf{c}} - W_{Z^{\mathbf{r}_m,\mathbf{c}_n}}\|_2 \to 0$ as $m, n \to \infty$.

Theorem (L., and Muhkerjee '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 *as* $m, n \to \infty$.

(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) *For* $C = C(\mu, \delta) > 0$,

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

In particular, $\| W^{\mathbf{r},\mathbf{c}} - W_{Z^{\mathbf{r}_m,\mathbf{c}_n}} \|_2 \to 0$ *as* $m, n \to \infty$.

(iii) *Denote* $s(m,n) := mn\sqrt{m^{-1} + n^{-1}}$ *and let* $X \sim \mu^{\otimes(m \times n)}$ *be conditional on* $X \in \mathcal{T}_{s(m,n)}(\mathbf{r}_m, \mathbf{c}_n)$. *Then with prob. at least* $1 - \exp(-C_1 s(m,n))$,

$$\| W_X - W^{\mathbf{r},\mathbf{c}} \|_\square \leq \underbrace{C_2 \left(m^{-1} + n^{-1} \right)^{1/4}}_{\text{fluctuation around } W_{Z^{\mathbf{r}_m,\mathbf{c}_n}}} + \underbrace{C_2 \sqrt{\| (\mathbf{r}, \mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n) \|}}_{\text{bias } (*)}$$

Corollary (L., and Muhkerjee '24+)

Assume $\mathrm{Supp}(\mu) = \{0, 1, \cdots, B\}$ *for some* $B \in \{1, 2, \cdots, \infty\}$.

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 *as* $m, n \to \infty$.

Let $X \sim \mu^{\otimes(m \times n)}$ *be conditioned on* ***satisfying the margin*** $(\mathbf{r}_m, \mathbf{c}_n)$ ***exactly***.

Corollary (L., and Muhkerjee '24+)

Assume $\mathrm{Supp}(\mu) = \{0, 1, \cdots, B\}$ *for some* $B \in \{1, 2, \cdots, \infty\}$.

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

Let $X \sim \mu^{\otimes(m \times n)}$ *be conditioned on* ***satisfying the margin*** $(\mathbf{r}_m, \mathbf{c}_n)$ ***exactly***.

(i) *(Finite support) Suppose* $B < \infty$. *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/4)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_{\square} \leq c_1 \, (mn)^{-(1/8)+\varepsilon} + c_2 \sqrt{\|(\mathbf{r}, \mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

Corollary (L., and Muhkerjee '24+)

Assume $\mathrm{Supp}(\mu) = \{0, 1, \cdots, B\}$ *for some* $B \in \{1, 2, \cdots, \infty\}$.

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 *as* $m, n \to \infty$.

Let $X \sim \mu^{\otimes(m\times n)}$ *be conditioned on* ***satisfying the margin*** $(\mathbf{r}_m, \mathbf{c}_n)$ ***exactly***.

(i) *(Finite support) Suppose* $B < \infty$. *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/4)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \le c_1\, (mn)^{-(1/8)+\varepsilon} + c_2 \sqrt{\|(\mathbf{r}, \mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

(ii) *(Infinite support) Suppose* μ *is the counting measure on* $\mathbb{Z}_{\ge 0}$ *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/2)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \le c_1\, (mn)^{-(1/4)+\varepsilon} + c_2 \sqrt{\|(\mathbf{r}, \mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

Corollary (L., and Muhkerjee '24+)

Assume $\mathrm{Supp}(\mu) = \{0, 1, \cdots, B\}$ *for some* $B \in \{1, 2, \cdots, \infty\}$.

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 *as* $m, n \to \infty$.

Let $X \sim \mu^{\otimes(m \times n)}$ *be conditioned on* ***satisfying the margin*** $(\mathbf{r}_m, \mathbf{c}_n)$ ***exactly***.

(i) *(Finite support) Suppose* $B < \infty$. *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/4)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_{\square} \le c_1\,(mn)^{-(1/8)+\varepsilon} + c_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

(ii) *(Infinite support) Suppose* μ *is the counting measure on* $\mathbb{Z}_{\geq 0}$ *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/2)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_{\square} \le c_1\,(mn)^{-(1/4)+\varepsilon} + c_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

- We can do exact conditioning whenever $\mathbb{P}(Y \in \mathcal{T}(\mathbf{r}_m, \mathbf{c}_n)) \geq \exp(-o(mn))$, where $Y \sim$ ML (α, β)-model

Corollary (L., and Muhkerjee '24+)

Assume $\mathrm{Supp}(\mu) = \{0, 1, \cdots, B\}$ *for some* $B \in \{1, 2, \cdots, \infty\}$.

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 *as* $m, n \to \infty$.

Let $X \sim \mu^{\otimes(m\times n)}$ *be conditioned on* ***satisfying the margin*** $(\mathbf{r}_m, \mathbf{c}_n)$ ***exactly***.

(i) *(Finite support) Suppose* $B < \infty$. *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/4)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \le c_1\,(mn)^{-(1/8)+\varepsilon} + c_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

(ii) *(Infinite support) Suppose* μ *is the counting measure on* $\mathbb{Z}_{\ge 0}$ *Then for each* $\varepsilon > 0$, *with probability at least* $1 - \exp\left(-(mn)^{(1/2)+\varepsilon}\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \le c_1\,(mn)^{-(1/4)+\varepsilon} + c_2\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

- We can do exact conditioning whenever $\mathbb{P}(Y \in \mathcal{T}(\mathbf{r}_m, \mathbf{c}_n)) \ge \exp(-o(mn))$, where $Y \sim$ ML (α, β)-model
- δ-tame condition $\Leftarrow$ 'asymmetric EG-condition' for **(i)**, 'subcriticality' for **(ii)**

Outline

- There is rich literature on the connection between Schrödinger bridge and optimal transport (see, e.g., [11, 13]).

- There is rich literature on the connection between Schrödinger bridge and optimal transport (see, e.g., [11, 13]).
  - For $\gamma : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ a **cost function** and $\varepsilon > 0$ fixed, then taking $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, then

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

which is the **optimal transport** with **entropic regularization** [14].

- There is rich literature on the connection between Schrödinger bridge and optimal transport (see, e.g., [11, 13]).
  - For $\gamma : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ a **cost function** and $\varepsilon > 0$ fixed, then taking $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, then

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

    which is the **optimal transport** with **entropic regularization** [14].
  - Classical case with quadratic cost: $\gamma(x,y) = (x-y)^2$

Open problems

- There is rich literature on the connection between Schrödinger bridge and optimal transport (see, e.g., [11, 13]).
  - For $\gamma : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ a **cost function** and $\varepsilon > 0$ fixed, then taking $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, then

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

  which is the **optimal transport** with **entropic regularization** [14].
  - Classical case with quadratic cost: $\gamma(x,y) = (x-y)^2$
- An OT version of our margin-conditioned random matrix problem?

Open problems

- There is rich literature on the connection between Schrödinger bridge and optimal transport (see, e.g., [11, 13]).
  - For $\gamma : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ a **cost function** and $\varepsilon > 0$ fixed, then taking $\mathcal{R} \propto e^{-\gamma/\varepsilon}\mu_1 \otimes \mu_2$, then

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

    which is the **optimal transport** with **entropic regularization** [14].
  - Classical case with quadratic cost: $\gamma(x,y) = (x-y)^2$
- An OT version of our margin-conditioned random matrix problem?
  - $\gamma : \mathbb{R}^{m\times n} \to \mathbb{R}_{\geq 0}$ a cost function on $m \times n$ real matrices

Open problems

- There is rich literature on the connection between Schrödinger bridge and optimal transport (see, e.g., [11, 13]).
  - For $\gamma : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ a **cost function** and $\varepsilon > 0$ fixed, then taking $\mathcal{R} \propto e^{-\gamma/\varepsilon} \mu_1 \otimes \mu_2$, then

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

    which is the **optimal transport** with **entropic regularization** [14].
  - Classical case with quadratic cost: $\gamma(x, y) = (x - y)^2$
- An OT version of our margin-conditioned random matrix problem?
  - $\gamma : \mathbb{R}^{m \times n} \to \mathbb{R}_{\geq 0}$ a cost function on $m \times n$ real matrices
  - Replacing the base model $\mathcal{R}$ with a probability measure proportional to $e^{-\gamma/\varepsilon}\mathcal{R}$,

$$\min_{\mathcal{H} \in \mathcal{P}^{m \times n}} \int_{\mathbb{R}^{m \times n}} \gamma(\mathbf{x}) \mathcal{H}(d\mathbf{x}) + \varepsilon\, D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) \quad \text{s.t.} \quad \mathbb{E}_{X \sim \mathcal{H}}[(r(X), c(X))] = (\mathbf{r}, \mathbf{c})$$

    (*Natural matrix-loss function γ? Spectral norm?*)

- Extend the theory for more general base measure for RM ensemble than i.i.d.?

Open problems

- 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)?

Open problems

- 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?

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

Open problems

- 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?

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

- Limit theory for Shrödinger bridge / Optimal transport?

Open problems

- 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?

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

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

Open problems

- 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?

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

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

- 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?

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

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

Open problems

- 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?

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

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

Open problems

- 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?

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

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

Open problems

- 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?

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

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

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

[3] 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.

[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] 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] 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.

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

[10] 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.

[11] Simone Di Marino and Augusto Gerolin. "An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm". In: *Journal of Scientific Computing* 85.2 (2020), p. 27.

[12] Hoi H Nguyen. "Random doubly stochastic matrices: the circular law". In: (2014).

[13] Marcel Nutz. "Introduction to entropic optimal transport". In: *Lecture notes, Columbia University* (2021).

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