Large random matrices with given margins

Hanbaek Lyu

University of Wisconsin - Madison

Based on joint work with Sumit Mukherjee (Columbia)

Apr. 3, 2025

- (*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

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

- *(Base model)* μ = 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 $\mathbf{x}$:

$$r(\mathbf{x}) := (r_1(\mathbf{x}), \ldots, r_m(\mathbf{x})); \quad r_i(\mathbf{x}) := x_{i1} + \cdots + x_{in} \qquad (\triangleright \textit{ row margin of } \mathbf{x})$$

$$c(\mathbf{x}) := (c_1(\mathbf{x}), \ldots, c_n(\mathbf{x})); \quad c_j(\mathbf{x}) := x_{1j} + \cdots + x_{mj} \qquad (\triangleright \textit{ column margin of } \mathbf{x})$$

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

- *(Base model)* μ = 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 $\mathbf{x}$:

$$r(\mathbf{x}) := (r_1(\mathbf{x}), \dots, r_m(\mathbf{x})); \quad r_i(\mathbf{x}) := x_{i1} + \cdots + x_{in} \qquad (\triangleright \textit{ row margin of } \mathbf{x})$$

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

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

- *(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?

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

- (*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?

- $X_{11} \overset{d}{=} ?$
- $\mathbb{E}[X] = ?$
- $X - \mathbb{E}[X] \overset{d}{=} ?$
- $X \overset{d}{=} ?$

- (*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?

  - $X_{11} \stackrel{d}{=}$?
  - $\mathbb{E}[X] =$?
  - $X - \mathbb{E}[X] \stackrel{d}{=}$?
  - $X \stackrel{d}{=}$?

- High-level answer:

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

- (*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?

  - $X_{11} \overset{d}{=} ?$
  - $\mathbb{E}[X] = ?$
  - $X - \mathbb{E}[X] \overset{d}{=} ?$
  - $X \overset{d}{=} ?$

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

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

- (*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?

  - $X_{11} \overset{d}{=} ?$
  - $\mathbb{E}[X] = ?$
  - $X - \mathbb{E}[X] \overset{d}{=} ?$
  - $X \overset{d}{=} ?$

- High-level answer:
  - *Maximum Liklihood Perspective (Parameteric)*: The maximum likelihood entrywise exponential tilting of the base model for margin $(\mathbf{r},\mathbf{c})$
  - *Minimum Relative Entropy Perspective (Non-parametric)*: The random matrix ensemble with minimum relative entropy from the base model constrained to have the 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?

  - $X_{11} \overset{d}{=}$?
  - $\mathbb{E}[X] =$?
  - $X - \mathbb{E}[X] \overset{d}{=}$?
  - $X \overset{d}{=}$?

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

  - *Minimum Relative Entropy Perspective (Non-parametric)*: The random matrix ensemble with minimum relative entropy from the base model constrained to have the expected margin $(\mathbf{r},\mathbf{c})$.
- These two approaches give the same answer! (strong duality)

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

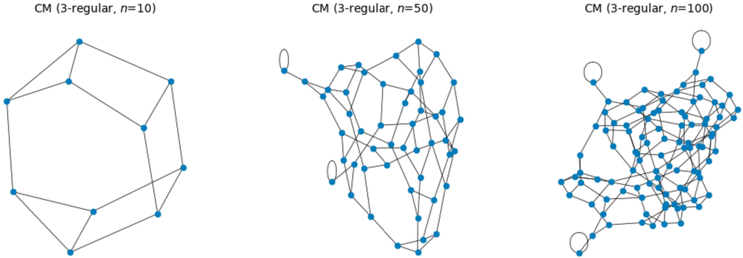


Figure: Random 3-regular graphs generated by the configuration model (allowing loops)

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

- $(\mathbf{d}^n)_{n\geq 1}$: dense degree sequence with scaling limit to $\mathbf{c} : [0,1] \to (c_1, c_2) \subseteq (0,1)$
- Chaterjee, Diaconis, Sly '11 [5]

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 [5]
  - Assume $\mathbf{c}$ satisfies the 'continuum Erdős-Gallai condition'

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

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 [5]
  - Assume $\mathbf{c}$ satisfies the 'continuum Erdős-Gallai condition'

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

  - 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 [5]
  - Assume $\mathbf{c}$ satisfies the 'continuum Erdős-Gallai condition'

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

  - 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

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

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

Key idea in CDS '11 [5]

- 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 [5]

- 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 $\boldsymbol{\beta} \in \mathbb{R}^n$, $G_{\boldsymbol{\beta}}$ = random graph with n nodes and independent edges, where

$$\mathbb{P}_{\boldsymbol{\beta}}(i \sim j) = \frac{e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}}{1 + e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}} = \mathbb{E}[A^{\boldsymbol{\beta}}(i,j)]$$

- 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 $\boldsymbol{\beta} \in \mathbb{R}^n$, $G_{\boldsymbol{\beta}}$ = random graph with n nodes and independent edges, where

$$\mathbb{P}_{\boldsymbol{\beta}}(i \sim j) = \frac{e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}}{1 + e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}} = \mathbb{E}[A^{\boldsymbol{\beta}}(i,j)]$$

  - (log-likelihood)

$$\ell(\boldsymbol{\beta}) = \sum_{i,j} x_{ij}(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j)) - \underbrace{\log\left(1 + e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}\right)}_{=\psi(\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j))} = 2\langle \mathbf{d}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\beta}(i) + \boldsymbol{\beta}(j))$$

Key idea in CDS '11 [5]

- 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 $\boldsymbol{\beta} \in \mathbb{R}^n$, $G_{\boldsymbol{\beta}}$ = random graph with n nodes and independent edges, where

$$\mathbb{P}_{\boldsymbol{\beta}}(i \sim j) = \frac{e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}}{1+e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}} = \mathbb{E}[A^{\boldsymbol{\beta}}(i,j)]$$

  - (log-likelihood)

$$\ell(\boldsymbol{\beta}) = \sum_{i,j} x_{ij}(\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)) - \underbrace{\log\left(1+e^{\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j)}\right)}_{=\psi(\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j))} = 2\langle \mathbf{d}, \boldsymbol{\beta}\rangle - \sum_{i,j}\psi(\boldsymbol{\beta}(i)+\boldsymbol{\beta}(j))$$

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

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

Key idea in CDS '11 [5]

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

Key idea in CDS '11 [5]

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

$$G^{\beta^n} \quad \overset{d}{\approx} \quad G(\mathbf{d}^n) \sim \text{Uniform}(G \text{ with deg seq } = \mathbf{d}^n)$$

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

$$G^{\beta^n} \quad \overset{d}{\approx} \quad G(\mathbf{d}^n) \sim \text{Uniform}(G \text{ with deg seq } = \mathbf{d}^n)$$

1. $G^{\beta^n} \mid \{\text{deg seq} = \mathbf{d}^n\} \quad \overset{d}{=} \quad G(\mathbf{d}^n)$

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

$$G^{\beta^n} \quad \overset{d}{\approx} \quad G(\mathbf{d}^n) \sim \text{Uniform}(G \text{ with deg seq } = \mathbf{d}^n)$$

1. $G^{\beta^n} \mid \{\text{deg seq} = \mathbf{d}^n\} \quad \overset{d}{=} \quad G(\mathbf{d}^n)$

2. $\mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n) \geq \exp(-o(n^{3/2+\varepsilon}))$

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

$$G^{\beta^n} \quad \overset{d}{\approx} \quad G(\mathbf{d}^n) \sim \text{Uniform}(G \text{ with deg seq } = \mathbf{d}^n)$$

1. $G^{\beta^n} \mid \{\text{deg seq} = \mathbf{d}^n\} \quad \overset{d}{=} \quad G(\mathbf{d}^n)$

2. $\mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n) \geq \exp(-o(n^{3/2+\varepsilon}))$

  - The above implies

$$\begin{aligned} \mathbb{P}\left(G(\mathbf{d}^n) \in \mathcal{E}\right) &= \mathbb{P}\left(G^{\beta^n} \in \mathcal{E} \,\middle|\, G^{\beta^n} \text{ has deg seq } \mathbf{d}^n\right) \\ &\leq \mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n)^{-1} \mathbb{P}\left(G^{\beta^n} \in \mathcal{E}\right) \\ &\leq \exp(o(n^{3/2+\varepsilon}))\, \mathbb{P}\left(G^{\beta^n} \in \mathcal{E}\right) \end{aligned}$$

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

$$G^{\beta^n} \quad \overset{d}{\approx} \quad G(\mathbf{d}^n) \sim \text{Uniform}(G \text{ with deg seq } = \mathbf{d}^n)$$

1. $G^{\beta^n} \mid \{\text{deg seq} = \mathbf{d}^n\} \quad \overset{d}{=} \quad G(\mathbf{d}^n)$

2. $\mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n) \geq \exp(-o(n^{3/2+\varepsilon}))$

  - The above implies

$$\begin{aligned}\mathbb{P}\left(G(\mathbf{d}^n) \in \mathcal{E}\right) &= \mathbb{P}\left(G^{\beta^n} \in \mathcal{E} \,\middle|\, G^{\beta^n} \text{ has deg seq } \mathbf{d}^n\right) \\ &\leq \mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n)^{-1}\mathbb{P}\left(G^{\beta^n} \in \mathcal{E}\right) \\ &\leq \exp(o(n^{3/2+\varepsilon}))\ \mathbb{P}\left(G^{\beta^n} \in \mathcal{E}\right)\end{aligned}$$

  - So events extremely rare under G^{β^n} are also rare under $G(\mathbf{d}^n)$

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

$$G^{\beta^n} \quad \overset{d}{\approx} \quad G(\mathbf{d}^n) \sim \text{Uniform}(G \text{ with deg seq } = \mathbf{d}^n)$$

1. $G^{\beta^n} \,|\, \{\text{deg seq} = \mathbf{d}^n\} \quad \overset{d}{=} \quad G(\mathbf{d}^n)$

2. $\mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n) \geq \exp(-o(n^{3/2+\varepsilon}))$

  - The above implies

$$\begin{aligned}\mathbb{P}\left(G(\mathbf{d}^n) \in \mathcal{E}\right) &= \mathbb{P}\left(G^{\beta^n} \in \mathcal{E} \,\middle|\, G^{\beta^n} \text{ has deg seq } \mathbf{d}^n\right) \\ &\leq \mathbb{P}(G^{\beta^n} \text{ has deg seq } \mathbf{d}^n)^{-1} \mathbb{P}\left(G^{\beta^n} \in \mathcal{E}\right) \\ &\leq \exp(o(n^{3/2+\varepsilon}))\, \mathbb{P}\left(G^{\beta^n} \in \mathcal{E}\right)\end{aligned}$$

  - So events extremely rare under G^{β^n} are also rare under $G(\mathbf{d}^n)$

    e.g., $G(\mathbf{d}^n)$ cannot be too far from $\mathbb{E}[G^{\beta^n}] \approx G^{\beta^n}$

Exponential tilting

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$

Exponential tilting

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})$
- $\mu_\theta :=$ exponentially tilted probability measure given by

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

- Goal: Approximate $X \sim \mu^{\otimes(m \times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- μ_θ := exponentially tilted probability measure given by

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

- Elementary facts:

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

  - ψ' : {tilting params.} $\to (A, B)$ is strictly increasing (▷ `tilt2mean` function)
  - $\phi = (\psi')^{-1} : (A, B) \to$ {tilting params.} is strictly increasing (▷ `mean2tilt` function)

- Goal: Approximate $X \sim \mu^{\otimes(m \times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$

The (α, β)-model

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- Given vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: RM w/ indep. entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$

The (α, β)-model

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- Given vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: RM w/ indep. entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$
- log-likelihood of margin $(\mathbf{r}, \mathbf{c})$ under Y w.r.t. $\mu^{\otimes(m\times n)}$:

$$\sum_{i,j}\Big[x_{ij}(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)) - \psi(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j))\Big] = \langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \mathbf{c}, \boldsymbol{\beta}\rangle - \sum_{i,j}\psi(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)).$$

- Goal: Approximate $X \sim \mu^{\otimes(m \times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- Given vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$: RM w/ indep. entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}$
- log-likelihood of margin $(\mathbf{r}, \mathbf{c})$ under Y w.r.t. $\mu^{\otimes(m \times n)}$:

$$\sum_{i,j} \left[x_{ij}(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) - \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right] = \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)).$$

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

$$\sup_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right),$$

The (α, β)-model

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- Given vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: RM w/ indep. entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$
- log-likelihood of margin $(\mathbf{r}, \mathbf{c})$ under Y w.r.t. $\mu^{\otimes(m\times n)}$:

$$\sum_{i,j} \Big[x_{ij}(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) - \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \Big] = \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)).$$

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

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right),$$

- Taking $\nabla g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{0}$, get the MLE equation:

$$\mathbb{E}[Y] = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \in \mathcal{T}(\mathbf{r}, \mathbf{c})$$

The (α, β)-model

- Goal: Approximate $X \sim \mu^{\otimes(m \times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- Given vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$: RM w/ indep. entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}$
- log-likelihood of margin $(\mathbf{r}, \mathbf{c})$ under Y w.r.t. $\mu^{\otimes(m \times n)}$:

$$\sum_{i,j} \left[x_{ij}(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) - \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right] = \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)).$$

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

$$\sup_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \left(g^{\mathbf{r}, \mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right),$$

- Taking $\nabla g^{\mathbf{r}, \mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{0}$, get the MLE equation:

$$\mathbb{E}[Y] = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \in \mathcal{T}(\mathbf{r}, \mathbf{c})$$

  - MLE is not unique: $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \iff (\boldsymbol{\alpha} + \lambda, \boldsymbol{\beta} - \lambda)$

The $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})$
- Given vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $Y = (Y_{ij}) \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$: RM w/ indep. entries $Y_{ij} \sim \mu_{\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)}$
- log-likelihood of margin $(\mathbf{r},\mathbf{c})$ under Y w.r.t. $\mu^{\otimes(m\times n)}$:

$$\sum_{i,j}\Big[x_{ij}(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)) - \psi(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j))\Big] = \langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \mathbf{c}, \boldsymbol{\beta}\rangle - \sum_{i,j}\psi(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j)).$$

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

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \mathbf{c}, \boldsymbol{\beta}\rangle - \sum_{i,j}\psi(\boldsymbol{\alpha}(i)+\boldsymbol{\beta}(j))\right),$$

- Taking $\nabla g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \mathbf{0}$, get the MLE equation:

$$\mathbb{E}[Y] = \psi'(\boldsymbol{\alpha}\oplus\boldsymbol{\beta}) \in \mathcal{T}(\mathbf{r},\mathbf{c})$$

  - MLE is not unique: $(\boldsymbol{\alpha},\boldsymbol{\beta}) \iff (\boldsymbol{\alpha}+\lambda, \boldsymbol{\beta}-\lambda)$
  - How do we compute an MLE? No closed form; use Sinkhorn-type algorithm (will revisit)

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

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

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

where $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is an MLE for $(\mathbf{r},\mathbf{c})$

- Behavior of an $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model depends crucially on how far the entries of $\boldsymbol{\alpha} \oplus \boldsymbol{\beta}$ are away from the extreme values $\phi(A)$ and $\phi(B)$

- Behavior of an $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model depends crucially on how far the entries of $\boldsymbol{\alpha} \oplus \boldsymbol{\beta}$ are away from the extreme values $\phi(A)$ and $\phi(B)$

Definition (Tame margins)

An $m \times n$ margin $(\mathbf{r}, \mathbf{c})$ is *δ-tame* for $\delta > 0$ if the MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy (recall $(A, B) = \mathrm{Int}(\mathrm{supp}(\mu))$)

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \leq \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \leq \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- Behavior of an $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model depends crucially on how far the entries of $\boldsymbol{\alpha} \oplus \boldsymbol{\beta}$ are away from the extreme values $\phi(A)$ and $\phi(B)$

Definition (Tame margins)

An $m \times n$ margin $(\mathbf{r}, \mathbf{c})$ is *δ-tame* for $\delta > 0$ if the MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy (recall $(A, B) = \mathrm{Int}(\mathrm{supp}(\mu))$)

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \leq \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \leq \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- If an MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists for $(\mathbf{r}, \mathbf{c})$, then $(\mathbf{r}, \mathbf{c})$ is always δ-tame for some $\delta > 0$ that may depend on m and n.

- Behavior of an $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-model depends crucially on how far the entries of $\boldsymbol{\alpha} \oplus \boldsymbol{\beta}$ are away from the extreme values $\phi(A)$ and $\phi(B)$

Definition (Tame margins)

An $m \times n$ margin $(\mathbf{r}, \mathbf{c})$ is *δ-tame* for $\delta > 0$ if the MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy (recall $(A, B) = \mathrm{Int}(\mathrm{supp}(\mu))$)

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \leq \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \leq \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- If an MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists for $(\mathbf{r}, \mathbf{c})$, then $(\mathbf{r}, \mathbf{c})$ is always δ-tame for some $\delta > 0$ that may depend on m and n.
- A technical issue: When is a sequence of margins **uniformly δ-tame**? (will revisit)

Theorem (Transference; L-M '24+)

$(\mathbf{r}, \mathbf{c}) = m \times n$ δ-*tame margin with an MLE* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, *and let* $X \sim \mu^{\otimes (m \times n)}$ *be conditional on* $X \in \mathcal{T}_{\rho}(\mathbf{r}, \mathbf{c})$ *for some* $\rho \geq 0$. *Let* $Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$. *Then for some constant* $C = C(\mu, \delta) > 0$, *for each measurable set* $\mathcal{E} \subseteq \mathbb{R}^{m \times n}$,

Theorem (Transference; L-M '24+)

$(\mathbf{r},\mathbf{c}) = m \times n$ *δ-tame margin with an MLE* $(\boldsymbol{\alpha},\boldsymbol{\beta})$*, and let* $X \sim \mu^{\otimes(m\times n)}$ *be conditional on* $X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})$ *for some* $\rho \geq 0$*. Let* $Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$*. Then for some constant* $C = C(\mu,\delta) > 0$*, for each measurable set* $\mathcal{E} \subseteq \mathbb{R}^{m\times n}$*,*

$$\mathbb{P}(X \in \mathcal{E}) \leq \exp\left(C\rho\right)\mathbb{P}\left(Y \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right)^{-1}\mathbb{P}\left(Y \in \mathcal{E}\right).$$

Theorem (Transference; L-M '24+)

$(\mathbf{r},\mathbf{c}) = m \times n$ δ-tame margin with an MLE $(\boldsymbol{\alpha},\boldsymbol{\beta})$, and let $X \sim \mu^{\otimes(m\times n)}$ be conditional on $X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})$ for some $\rho \geq 0$. Let $Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$. Then for some constant $C = C(\mu,\delta) > 0$, for each measurable set $\mathcal{E} \subseteq \mathbb{R}^{m\times n}$,

$$\mathbb{P}(X \in \mathcal{E}) \leq \exp\left(C\rho\right)\mathbb{P}\left(Y \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right)^{-1}\mathbb{P}\left(Y \in \mathcal{E}\right).$$

In particular, if $\rho \geq \sqrt{mn(m+n)}$, then for each $t \geq 0$,

$$\mathbb{P}(X \in \mathcal{E}) \leq 2\exp\left(C\rho\right)\mathbb{P}\left(Y \in \mathcal{E}\right).$$

Theorem (Transference; L-M '24+)

$(\mathbf{r},\mathbf{c}) = m\times n$ δ-tame margin with an MLE $(\boldsymbol{\alpha},\boldsymbol{\beta})$, and let $X\sim\mu^{\otimes(m\times n)}$ be conditional on $X\in\mathcal{T}_\rho(\mathbf{r},\mathbf{c})$ for some $\rho\geq 0$. Let $Y\sim\boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$. Then for some constant $C=C(\mu,\delta)>0$, for each measurable set $\mathcal{E}\subseteq\mathbb{R}^{m\times n}$,

$$\mathbb{P}(X\in\mathcal{E}) \leq \exp\left(C\rho\right)\mathbb{P}\left(Y\in\mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right)^{-1}\mathbb{P}\left(Y\in\mathcal{E}\right).$$

In particular, if $\rho\geq\sqrt{mn(m+n)}$, then for each $t\geq 0$,

$$\mathbb{P}(X\in\mathcal{E}) \leq 2\exp\left(C\rho\right)\mathbb{P}\left(Y\in\mathcal{E}\right).$$

- Events extremely rare under Y are also rare under X

Theorem (Transference; L-M '24+)

$(\mathbf{r}, \mathbf{c}) = m \times n$ δ-tame margin with an MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and let $X \sim \mu^{\otimes(m\times n)}$ be conditional on $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$ for some $\rho \geq 0$. Let $Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$. Then for some constant $C = C(\mu, \delta) > 0$, for each measurable set $\mathcal{E} \subseteq \mathbb{R}^{m\times n}$,

$$\mathbb{P}(X \in \mathcal{E}) \leq \exp(C\rho)\, \mathbb{P}(Y \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c}))^{-1}\, \mathbb{P}(Y \in \mathcal{E}).$$

In particular, if $\rho \geq \sqrt{mn(m+n)}$, then for each $t \geq 0$,

$$\mathbb{P}(X \in \mathcal{E}) \leq 2\exp(C\rho)\, \mathbb{P}(Y \in \mathcal{E}).$$

- Events extremely rare under Y are also rare under X
- $\mathbb{P}(Y \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})) \geq 1/2$ for $\rho \sim \sqrt{mn(m+n)}$ $(\mathbb{E}[Y] \in \mathcal{T}(\mathbf{r}, \mathbf{c})$ and use concentration for $Y)$

- A *kernel* is an integrable function $W : [0,1]^2 \to \mathbb{R}$. The *cut-norm* of a kernel W is defined as

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

Given an $m \times n$ matrix A, define a step-kernel W_A as

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

- A *kernel* is an integrable function $W\colon [0,1]^2 \to \mathbb{R}$. The *cut-norm* of a kernel W is defined as

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

Given an $m \times n$ matrix A, define a step-kernel W_A as

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

Theorem (Concentration in cut norm; L-M '24+)

Keep the same setting as before. Then there exists a constant $C = C(\delta,\mu) > 0$ s.t.

$$\mathbb{P}\left(\|W_X - W_{\mathbb{E}[Y]}\|_\square \geq t\right) \leq \underbrace{\mathbb{P}\left(Y \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right)^{-1}}_{\text{transference cost}} \underbrace{\exp\left(C\rho + (m+n+1)\log 2 - \frac{t^2 mn}{C}\right)}_{\text{Concentration of } \|Y-\mathbb{E}[Y]\|_\square}.$$

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"

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

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

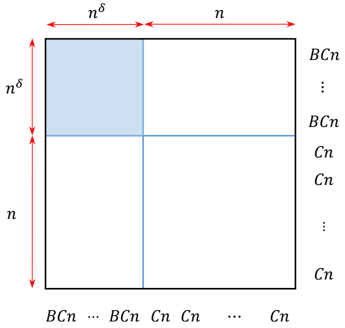
- Given implicitly by Good in 1963 [13] and later formally in 1963 [11] and 1976 [12]
- Experimentally verified by Good and Crook [10] in 1977 and Diagonis and Gangolli [7] in 1995
- Canfield and McKay '10 [4]: For $m = n$ and $\mathbf{r} = \mathbf{c} = (\lfloor Cn \rfloor, \dots, \lfloor Cn \rfloor)$,

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

- In 2008, Greenhill and McKay [14] proved same asymptotics for uniform but sparse margins: $\max(\mathbf{r})\cdot\max(\mathbf{c}) = O(N^{2/3})$

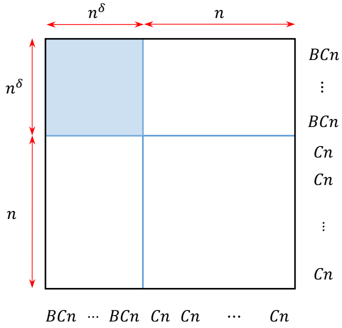
But what about non-uniform margins?

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



But what about non-uniform margins?

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



- 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

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

Maximum entropy method and 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 [2]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,max}_{Q \in \mathcal{T}(\mathbf{r},\mathbf{c})} \left[g(Z) = \sum_{i,j} \underbrace{(z_{ij}+1)\log(z_{ij}+1) - z_{ij}\log(z_{ij})}_{=\text{Entropy}(\text{Geom}(\text{mean}=z_{ij}))} \right]$$

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* Uniform$(\mathcal{T}(\mathbf{r},\mathbf{c}))$"
- (Barvinok '10 [2]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,max}_{Q \in \mathcal{T}(\mathbf{r},\mathbf{c})} \left[g(Z) = \sum_{i,j} \underbrace{(z_{ij}+1)\log(z_{ij}+1) - z_{ij}\log(z_{ij})}_{=\text{Entropy}(\text{Geom}(\text{mean} = z_{ij}))} \right]$$

- Barvinok's insight:

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

Maximum entropy method and 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 [2]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,max}_{Q \in \mathcal{T}(\mathbf{r},\mathbf{c})} \left[g(Z) = \sum_{i,j} \underbrace{(z_{ij}+1)\log(z_{ij}+1) - z_{ij}\log(z_{ij})}_{=\text{Entropy}(\text{Geom}(\text{mean}\,=\,z_{ij}))} \right]$$

- Barvinok's insight:

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

  - (Barvinok '09 [1], '10 [2])

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

- Barvinok '10: "*To count* $|\mathcal{T}(\mathbf{r},\mathbf{c})|$, *need to better understand* $\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c}))$"
- (Barvinok '10 [2]) For a $m \times n$ margin $(\mathbf{r},\mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \operatorname*{arg\,max}_{Q \in \mathcal{T}(\mathbf{r},\mathbf{c})} \left[g(Z) = \sum_{i,j} \underbrace{(z_{ij}+1)\log(z_{ij}+1) - z_{ij}\log(z_{ij})}_{=\text{Entropy}(\text{Geom}(\text{mean} = z_{ij}))} \right]$$

- Barvinok's insight:

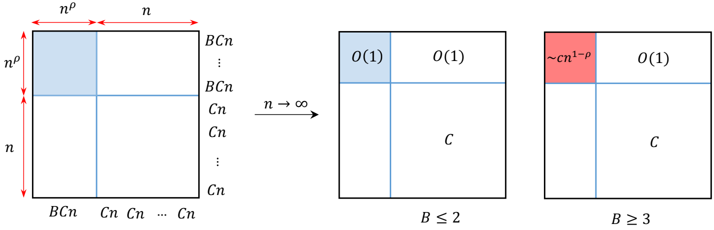
$$\text{Uniform}(\mathcal{T}(\mathbf{r},\mathbf{c})) \approx Z^{\mathbf{r},\mathbf{c}}$$

  - (Barvinok '09 [1], '10 [2])

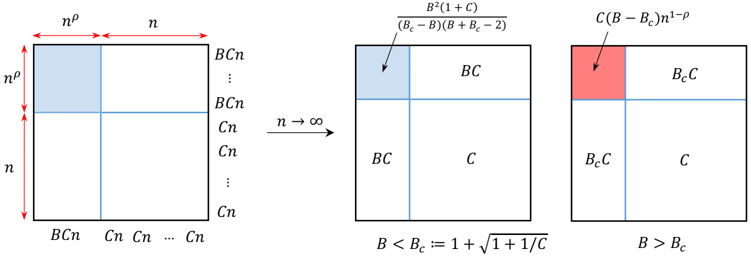
$$g(Z^{\mathbf{r},\mathbf{c}}) - \gamma n \log n \le \log|\mathcal{T}(\mathbf{r},\mathbf{c})| \le g(Z^{\mathbf{r},\mathbf{c}})$$

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

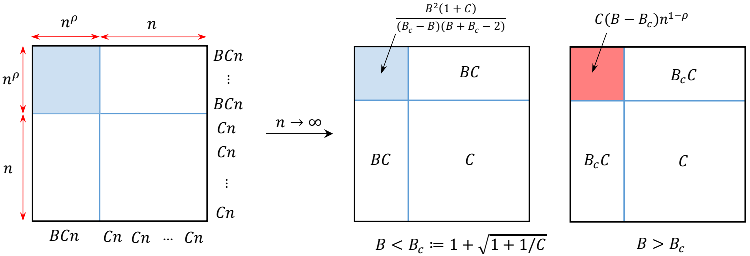
- In 2010, Barbinok conjectured that there is a phase transition in Uniform($\mathcal{T}$(Barv. margin)) as B increases



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

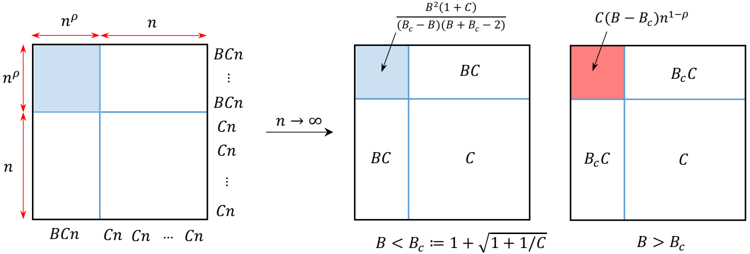


- Typical tables can change drastically by a small change in the margin!
  - For $0 \leq \delta < 1$, Dittmer, Lyu, and Pak [8] show that $Z^{\mathbf{r},\mathbf{c}}$ undergoes a **sharp phase transition** at $B_c = 1 + \sqrt{1 + C^{-1}}$:

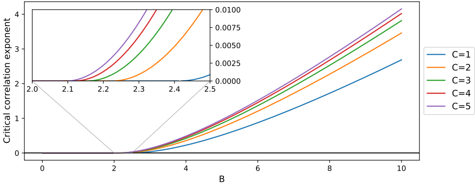


  - δ-tame for $B < B_c$, non-tame for $B > B_c$

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



  - δ-tame for $B < B_c$, non-tame for $B > B_c$
  - This result was used to obtain a second-order phase transition in the number of CTs with Barvinok margin by Lyu and Pak '22 [15] $(\log |\mathcal{T}(\mathbf{r}, \mathbf{c})| \approx g(Z^{\mathbf{r},\mathbf{c}}))$



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

Minimum relative entropy model

- Goal: Approximate $X \sim \mu^{\otimes(m \times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$

Minimum relative entropy model

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r}, \mathbf{c})$
- For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of $\mathcal{H}$ from $\mathcal{R}$* is

$$D_{KL}(\mathcal{H} \,\|\, \mathcal{R}) := \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \mathcal{R}(d\mathbf{x}) \quad \textit{if } \mathcal{H} \ll \mathcal{R} \textit{ and } \infty \textit{ o/w},$$

Minimum relative entropy model

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})$
- For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of* $\mathcal{H}$ *from* $\mathcal{R}$ is

$$D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) := \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \mathcal{R}(d\mathbf{x}) \quad \textit{if } \mathcal{H} \ll \mathcal{R} \textit{ and } \infty \textit{ o/w},$$

- Minimum relative entropy principle (information projection)

$$\left[X \sim \mu^{\otimes(m\times n)} \text{ given } X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right]$$
$$\overset{d}{\approx} \operatorname*{arg\,min}_{\mathcal{H}\in\mathcal{P}^{m\times n}} D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) \quad \text{subject to} \quad \mathbb{E}_{X\sim\mathcal{H}}[(r(X), c(X))] = (\mathbf{r},\mathbf{c})$$

- Goal: Approximate $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})$
- For each probability measure $\mathcal{H}$ on $\mathbb{R}^{m\times n}$, the *relative entropy of* $\mathcal{H}$ *from* $\mathcal{R}$ is

$$D_{KL}(\mathcal{H}\,\|\,\mathcal{R}) := \int_{\mathbf{x}\in\mathbb{R}^{m\times n}} \frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})} \log\left(\frac{d\mathcal{H}(\mathbf{x})}{d\mathcal{R}(\mathbf{x})}\right) \mathcal{R}(d\mathbf{x}) \quad \textit{if } \mathcal{H} \ll \mathcal{R} \textit{ and } \infty \textit{ o/w},$$

- Minimum relative entropy principle (information projection)

$$\left[X \sim \mu^{\otimes(m\times n)} \text{ given } X \in \mathcal{T}_\rho(\mathbf{r},\mathbf{c})\right]$$

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

$$= \bigotimes_{i,j} \mu_{\phi(z_{ij})} \quad \text{where} \quad Z = \operatorname*{arg\,min}_{Q=(q_{ij})\in\mathcal{T}(\mathbf{r},\mathbf{c})} \underbrace{D_{KL}\left(\bigotimes_{i,j}\mu_{\phi(q_{ij})} \,\middle\|\, \mu^{\otimes(m\times n)}\right)}_{=\sum_{i,j} D_{KL}(\mu_{\phi(q_{ij})}\|\mu)}$$

- The **relative entropy** from the base mesure μ to the tilted probability measure μ_θ:

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

Typical table

- The **relative entropy** from the base mesure μ to the tilted probability measure μ_θ:

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

- 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

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

  - 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

Typical table

- The **relative entropy** from the base mesure μ to the tilted probability measure μ_θ:

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

- 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

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

  - 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.

$$Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \qquad = \mathbb{E}[\boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}]!!$$

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

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

(*Informal result II: Minimum relative entropy perspective*)

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

where $Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta})$ **for some** $\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n$

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

- μ = Gaussian

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

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

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

- μ = Poisson

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

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

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

- $\mu = \text{Bernoulli}(1/2)$

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

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

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

Examples

- $\mu = \text{Bernoulli}(1/2)$

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

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

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

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

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

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

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

Dual(Typical Table)=MLE

Theorem (Strong duality; L-M '24+)

$$\arg\min_{Z} \ \underline{\texttt{Relative Entropy}}(Z) = \arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ \underline{\texttt{Likelihood Of Margin}}(\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$\arg\min_{Z\in\mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{i,j} D(\mu_{\phi(z_{ij})} \,\|\, \mu) = \arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left(\langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \sum_{i,j} \psi(\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)) \right)$$

$$\underbrace{Z^{\mathbf{r},\mathbf{c}}_{ij}}_{\text{typical table}} = \psi' \left(\underbrace{\boldsymbol{\alpha}(i) + \boldsymbol{\beta}(j)}_{=\text{MLE}} \right) \qquad (\psi' = \texttt{tilt2mean} \text{ function})$$

Static Schrödinger Bridge b/w $\mathbf{r}, \mathbf{c}$ = *Kantorovich dual with potential* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Non-parametric = Parametric

Scaling limit

- We know $X \approx Z^{\mathbf{r}_m, \mathbf{c}_n}$ in $\|\cdot\|_{\square}$. Does the typical tables (and the MLEs) have scaling limit as $(\mathbf{r}_m, \mathbf{c}_n) \to (\mathbf{r}, \mathbf{c})$?

Scaling limit

- We know $X \approx Z^{\mathbf{r}_m, \mathbf{c}_n}$ in $\| \cdot \|_\square$. Does the typical tables (and the MLEs) have scaling limit as $(\mathbf{r}_m, \mathbf{c}_n) \to (\mathbf{r}, \mathbf{c})$?
- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0,1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$

Scaling limit

- We know $X \approx Z^{\mathbf{r}_m, \mathbf{c}_n}$ in $\|\cdot\|_\square$. Does the typical tables (and the MLEs) have scaling limit as $(\mathbf{r}_m, \mathbf{c}_n) \to (\mathbf{r}, \mathbf{c})$?
- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0,1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$
- 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

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

- We know $X \approx Z^{\mathbf{r}_m,\mathbf{c}_n}$ in $\|\cdot\|_\square$. Does the typical tables (and the MLEs) have scaling limit as $(\mathbf{r}_m,\mathbf{c}_n) \to (\mathbf{r},\mathbf{c})$?
- A **continuum margin** $(\mathbf{r},\mathbf{c})$ = integrable functions $\mathbf{r},\mathbf{c} : (0,1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\,dx = \int_0^1 \mathbf{c}(y)\,dy$
- 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 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

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

Scaling limit

- We know $X \approx Z^{\mathbf{r}_m,\mathbf{c}_n}$ in $\|\cdot\|_\square$. Does the typical tables (and the MLEs) have scaling limit as $(\mathbf{r}_m, \mathbf{c}_n) \to (\mathbf{r}, \mathbf{c})$?
- A **continuum margin** $(\mathbf{r}, \mathbf{c})$ = integrable functions $\mathbf{r}, \mathbf{c} : (0, 1] \to \mathbb{R}$ such that $\int_0^1 \mathbf{r}(x)\, dx = \int_0^1 \mathbf{c}(y)\, dy$
- 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 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 $\mathcal{T}(\mathbf{r}_m, \mathbf{c}_n)$, $W_X \to W^{\mathbf{r},\mathbf{c}}$ a.s. in cut norm

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

Theorem (Scaling limit of typical tables and MLEs; L-M '24+)

Fix $\delta > 0$ *and let* $(\mathbf{r}_m, \mathbf{c}_n)$ *be a sequence of* $m \times n$ δ*-tame margins converging to a continuum margin* $(\mathbf{r}, \mathbf{c})$ *in* L^1 *as* $m, n \to \infty$*. Then* $\exists$ *bounded measurable functions* $\boldsymbol{\alpha}, \boldsymbol{\beta} : [0,1] \to \mathbb{R}$ *s.t.* $\int \boldsymbol{\alpha}(x) dx = 0$ *and the kernel*

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

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

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

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

Marginal distribution

- Recall the transference principle for exact margin conditioning ($\rho = 0$):

$$\mathbb{P}(X \in \mathcal{E}) \leq \mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})\right)^{-1} \mathbb{P}\left(Y \in \mathcal{E}\right), \quad Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$$

Marginal distribution

- Recall the transference principle for exact margin conditioning ($\rho = 0$):

$$\mathbb{P}(X \in \mathcal{E}) \leq \mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})\right)^{-1} \mathbb{P}\left(Y \in \mathcal{E}\right), \quad Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$$

- The key issue is to lower bound $\mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})\right)$

- Recall the transference principle for exact margin conditioning ($\rho = 0$):

$$\mathbb{P}(X \in \mathcal{E}) \leq \mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})\right)^{-1} \mathbb{P}\left(Y \in \mathcal{E}\right), \quad Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$$

- The key issue is to lower bound $\mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r}, \mathbf{c})\right)$

Corollary (L-M '24+)

Let $X \sim \mu^{\otimes(m \times n)}$ cond. on $X \in \mathcal{T}(\mathbf{r}, \mathbf{c})$. Suppose $(\mathbf{r}, \mathbf{c})$ is a 'k-cloning' of some $m_0 \times n_0$ margin $(\mathbf{a}, \mathbf{b})$ with an MLE $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, i.e., $\mathbf{r} = \mathbf{a} \otimes \mathbf{1}_k$ and $\mathbf{c} = \mathbf{b} \otimes \mathbf{1}_k$, where $\otimes$ denotes the Kronecker product. Then as $k \to \infty$,

$$d_{TV}(X_{11}, \mu_{\boldsymbol{\alpha}_0(1) + \boldsymbol{\beta}_0(1)}) = \begin{cases} O\left(k^{-1/2}\sqrt{\log k}\right) & \text{if } \mu = \text{Counting}(\mathbb{Z}_{\geq 0}) \text{ or Leb}(\mathbb{R}_{\geq 0}) \\ O\left(k^{-1/4}\log k\right) & \text{if } \mu = \text{Wtd versions of the above} \end{cases}$$

- Recall the transference principle for exact margin conditioning ($\rho = 0$):

$$\mathbb{P}(X \in \mathcal{E}) \leq \mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r},\mathbf{c})\right)^{-1} \mathbb{P}\left(Y \in \mathcal{E}\right), \quad Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha} \oplus \boldsymbol{\beta}}$$

- The key issue is to lower bound $\mathbb{P}\left(Y \in \mathcal{T}(\mathbf{r},\mathbf{c})\right)$

Corollary (L-M '24+)

Let $X \sim \mu^{\otimes(m\times n)}$ cond. on $X \in \mathcal{T}(\mathbf{r},\mathbf{c})$. Suppose $(\mathbf{r},\mathbf{c})$ is a 'k-cloning' of some $m_0 \times n_0$ margin $(\mathbf{a},\mathbf{b})$ with an MLE $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, i.e., $\mathbf{r} = \mathbf{a} \otimes \mathbf{1}_k$ and $\mathbf{c} = \mathbf{b} \otimes \mathbf{1}_k$, where $\otimes$ denotes the Kronecker product. Then as $k \to \infty$,

$$d_{TV}(X_{11}, \mu_{\boldsymbol{\alpha}_0(1)+\boldsymbol{\beta}_0(1)}) = \begin{cases} O\left(k^{-1/2}\sqrt{\log k}\right) & \text{if } \mu = \text{Counting}(\mathbb{Z}_{\geq 0}) \text{ or Leb}(\mathbb{R}_{\geq 0}) \\ O\left(k^{-1/4}\log k\right) & \text{if } \mu = \text{Wtd versions of the above} \end{cases}$$

- Answers Barvinok's 2010 conjecture on marginal distribution of random contingency tables

Theorem (Scaling limit in cut norm; L-M '24+)

Let δ*-tame margins* $(\mathbf{r}_m, \mathbf{c}_n) \to (\mathbf{r}, \mathbf{c})$ *in* L^1*. Let* $X \sim \mu^{\otimes(m\times n)}$ *cond. on* $X \in \mathcal{T}(\mathbf{r}_m, \mathbf{c}_n)$*.*
If $d\mu = h\, d\mathrm{Counting}(x)$ *or* $h\, dx$ *for "nice"* h*, with probability at least*
$1 - \exp\left(-C(m\sqrt{n} - n\sqrt{m})(\log(m+n))^2\right)$,

$$\|W_X - W^{\mathbf{r},\mathbf{c}}\|_\square \le C\sqrt{n^{-1/2} + m^{-1/2}}\log(m+n) + C\sqrt{\|(\mathbf{r},\mathbf{c}) - (\bar{\mathbf{r}}_m, \bar{\mathbf{c}}_n)\|_1}.$$

- We now know $X \approx \mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}}$. What about the fluctuation $X - Z^{\mathbf{r},\mathbf{c}}$?

- We now know $X \approx \mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}}$. What about the fluctuation $X - Z^{\mathbf{r},\mathbf{c}}$?
- Since X is a non-symmetric rectangluar RM, we look at the empirical singular value distribution (ESD).

- We now know $X \approx \mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}}$. What about the fluctuation $X - Z^{\mathbf{r},\mathbf{c}}$?
- Since X is a non-symmetric rectangluar RM, we look at the empirical singular value distribution (ESD).
  - If X is unconditioned, then the ESD follows Marchenko-Pastur quarter-circle law (1967). What happends under margin-conditioning?

- We now know $X \approx \mathbb{E}[Y] = Z^{\mathbf{r},\mathbf{c}}$. What about the fluctuation $X - Z^{\mathbf{r},\mathbf{c}}$?
- Since X is a non-symmetric rectangluar RM, we look at the empirical singular value distribution (ESD).
  - If X is unconditioned, then the ESD follows Marchenko-Pastur quarter-circle law (1967). What happends under margin-conditioning?

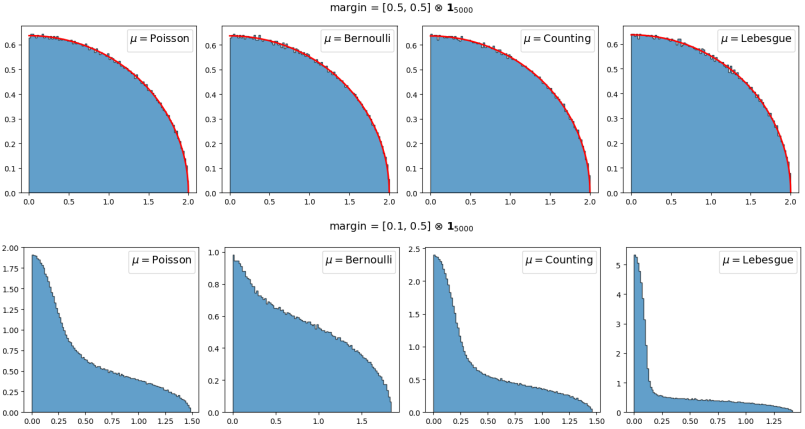


Figure: Empirical singular value distribution for $X \sim \mu^{\otimes(m\times n)}$ given $X \in \mathcal{T}(\mathbf{r}, \mathbf{c})$

Corollary (Universality of quarter-circle law for uniform margins; L-M 24+)

Assume uniform margins $\mathbf{r} = \mathbf{c} = a\mathbf{1}_n$ *for some* a *in the support of* μ*. Let*

$$\widetilde{X}_n := \frac{1}{\sqrt{2\psi''(\phi(a))n}}(X - a\mathbf{1}\mathbf{1}^\top).$$

Then the empirical singular value distribution of $\tilde{X}_n$ *converges weakly to the Marchenko-Pastur quarter-circle law* $\frac{1}{\pi}\sqrt{4-x^2}\,dx$ *in probability.*

Corollary (Universality of quarter-circle law for uniform margins; L-M 24+)

Assume uniform margins $\mathbf{r} = \mathbf{c} = a\mathbf{1}_n$ *for some* a *in the support of* μ*. Let*

$$\widetilde{X}_n := \frac{1}{\sqrt{2\psi''(\phi(a))n}}(X - a\mathbf{1}\mathbf{1}^\top).$$

Then the empirical singular value distribution of $\tilde{X}_n$ *converges weakly to the Marchenko-Pastur quarter-circle law* $\frac{1}{\pi}\sqrt{4-x^2}\,dx$ *in probability.*

- Chaterjee, Diaconis, Sly in 2010 showed the above for $\mu = \text{Leb}(\mathbb{R}_{\geq 0})$.

Corollary (Universality of quarter-circle law for uniform margins; L-M 24+)

Assume uniform margins $\mathbf{r} = \mathbf{c} = a\mathbf{1}_n$ *for some* a *in the support of* μ. *Let*

$$\widetilde{X}_n := \frac{1}{\sqrt{2\psi''(\phi(a))n}}(X - a\mathbf{1}\mathbf{1}^\top).$$

Then the empirical singular value distribution of $\tilde{X}_n$ *converges weakly to the Marchenko-Pastur quarter-circle law* $\frac{1}{\pi}\sqrt{4-x^2}\,dx$ *in probability.*

- Chaterjee, Diaconis, Sly in 2010 showed the above for $\mu = \mathrm{Leb}(\mathbb{R}_{\geq 0})$.
- Our result confirms universality of the M-P law for constant margin-conditoined RMs.

Corollary (Universality of quarter-circle law for uniform margins; L-M 24+)

Assume uniform margins $\mathbf{r} = \mathbf{c} = a\mathbf{1}_n$ *for some* a *in the support of* μ. *Let*

$$\widetilde{X}_n := \frac{1}{\sqrt{2\psi''(\phi(a))n}}(X - a\mathbf{1}\mathbf{1}^\top).$$

Then the empirical singular value distribution of $\tilde{X}_n$ *converges weakly to the Marchenko-Pastur quarter-circle law* $\frac{1}{\pi}\sqrt{4-x^2}\,dx$ *in probability.*

- Chaterjee, Diaconis, Sly in 2010 showed the above for $\mu = \mathrm{Leb}(\mathbb{R}_{\geq 0})$.
- Our result confirms universality of the M-P law for constant margin-conditoined RMs.
- We have a general result for arbitrary δ-tame marings. The limiting law is **not** always M-P; determined by the variance profile $\psi''(\boldsymbol{\alpha} \oplus \boldsymbol{\beta})$ through QVE determmining the Stieltjes transform.

Corollary (Universality of quarter-circle law for uniform margins; L-M 24+)

Assume uniform margins $\mathbf{r} = \mathbf{c} = a\mathbf{1}_n$ *for some* a *in the support of* μ. *Let*

$$\widetilde{X}_n := \frac{1}{\sqrt{2\psi''(\phi(a))n}}(X - a\mathbf{1}\mathbf{1}^\top).$$

Then the empirical singular value distribution of $\tilde{X}_n$ *converges weakly to the Marchenko-Pastur quarter-circle law* $\frac{1}{\pi}\sqrt{4 - x^2}\, dx$ *in probability.*

- Chaterjee, Diaconis, Sly in 2010 showed the above for $\mu = \mathrm{Leb}(\mathbb{R}_{\geq 0})$.
- Our result confirms universality of the M-P law for constant margin-conditoined RMs.
- We have a general result for arbitrary δ-tame marings. The limiting law is **not** always M-P; determined by the variance profile $\psi''(\boldsymbol{\alpha} \oplus \boldsymbol{\beta})$ through QVE determmining the Stieltjes transform.
- Sketch of Proof: $\mathrm{ESD}(\tilde{X}_n) \approx \mathrm{ESD}(\tilde{Y}_n)$ by transference; $\tilde{Y}_n \tilde{Y}_n^*$ generalized Wishart with variance profile $\psi''(\boldsymbol{\alpha}_n \oplus \boldsymbol{\beta}_n)$.

- All our results so far depends on the margin $(\mathbf{r}, \mathbf{c})$ being δ-tame: i.e., MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \leq \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \leq \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta .$$

- All our results so far depends on the margin $(\mathbf{r}, \mathbf{c})$ being δ-tame: i.e., MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \le \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \le \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- Tameness is an implicit condition. Seek for explicit conditions only using extreme values in the margin.

- All our results so far depends on the margin $(\mathbf{r}, \mathbf{c})$ being δ-tame: i.e., MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \le \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \le \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- Tameness is an implicit condition. Seek for explicit conditions only using extreme values in the margin.
- For each point $(s, t) \in (A, B)^2$, $s \le t$, we ask if an arbitrary $m \times n$ margin $(\mathbf{r}, \mathbf{c})$ satisftying

$$s \le \mathbf{r}(i)/n, \mathbf{c}(j)/m \le t \quad \text{for all } (i, j) \in [m] \times [n]$$

is δ-tame for some $\delta = \delta(\mu, s, t) > 0$ depending only on μ, s, and t.

- All our results so far depends on the margin $(\mathbf{r}, \mathbf{c})$ being δ-tame: i.e., MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \leq \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \leq \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- Tameness is an implicit condition. Seek for explicit conditions only using extreme values in the margin.
- For each point $(s, t) \in (A, B)^2$, $s \leq t$, we ask if an arbitrary $m \times n$ margin $(\mathbf{r}, \mathbf{c})$ satisftying

$$s \leq \mathbf{r}(i)/n, \mathbf{c}(j)/m \leq t \quad \text{for all } (i, j) \in [m] \times [n]$$

is δ-tame for some $\delta = \delta(\mu, s, t) > 0$ depending only on μ, s, and t.
- Let $\Omega(\mu) \subseteq (A, B)^2$ denote the set of all such points (s, t).

- All our results so far depends on the margin $(\mathbf{r}, \mathbf{c})$ being δ-tame: i.e., MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ exists and its entries satisfy

$$A_\delta := \max\left(A + \delta, -\frac{1}{\delta}\right) \leq \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \leq \min\left(B - \delta, \frac{1}{\delta}\right) =: B_\delta.$$

- Tameness is an implicit condition. Seek for explicit conditions only using extreme values in the margin.
- For each point $(s, t) \in (A, B)^2$, $s \leq t$, we ask if an arbitrary $m \times n$ margin $(\mathbf{r}, \mathbf{c})$ satisftying

$$s \leq \mathbf{r}(i)/n, \mathbf{c}(j)/m \leq t \quad \text{for all } (i, j) \in [m] \times [n]$$

  is δ-tame for some $\delta = \delta(\mu, s, t) > 0$ depending only on μ, s, and t.
- Let $\Omega(\mu) \subseteq (A, B)^2$ denote the set of all such points (s, t).
- Can we obtain the full phase diagram $\Omega(\mu)$ for each base measure μ?

Geometry of tame margins

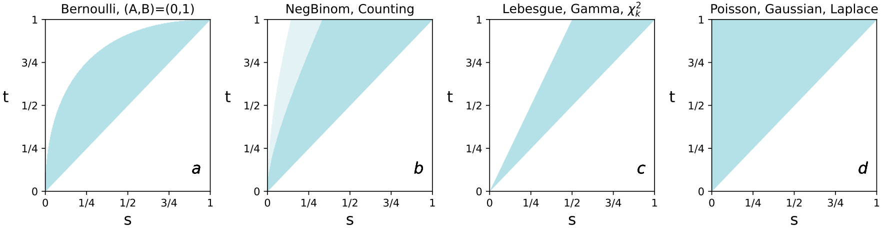


Figure: Phase diagrams for tame margins for various base measures μ. The upper contours are given by $(s+t)^2 < 2s$, $t \leq 1 + \sqrt{1 + rs^{-1}}$ ($r = 5$ for NegBinom and $r = 1$ for Counting), $t \leq s/2$, and $t = \infty$ from left to right.

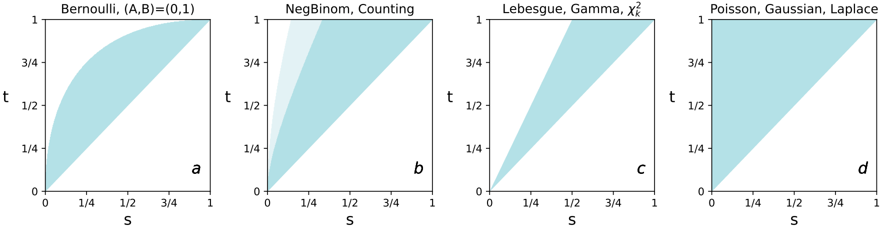


Figure: Phase diagrams for tame margins for various base measures μ. The upper contours are given by $(s+t)^2 < 2s$, $t \leq 1 + \sqrt{1 + rs^{-1}}$ ($r = 5$ for NegBinom and $r = 1$ for Counting), $t \leq s/2$, and $t = \infty$ from left to right.

Theorem (L-M '24+)

Suppose $-\infty < A \leq B < \infty$. Then each $(s, t) \in (A, B)^2$ with $s \leq t$ belongs to $\Omega(\mu)$ if

$$(s + t - 2A)^2 < 4(B - A)(s - A).$$

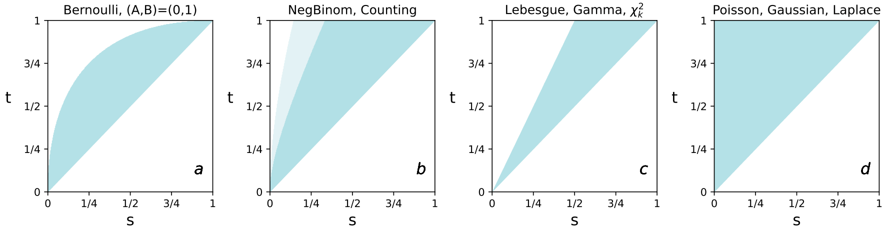


Figure: Phase diagrams for tame margins for various base measures μ. The upper contours are given by $(s+t)^2 < 2s$, $t \leq 1 + \sqrt{1 + rs^{-1}}$ ($r = 5$ for NegBinom and $r = 1$ for Counting), $t \leq s/2$, and $t = \infty$ from left to right.

Theorem (L-M '24+)

Suppose $-\infty < A \leq B < \infty$. *Then each* $(s, t) \in (A, B)^2$ *with* $s \leq t$ *belongs to* $\Omega(\mu)$ *if*

$$(s + t - 2A)^2 < 4(B - A)(s - A).$$

- A Matrix version of Erdös-Galai condition

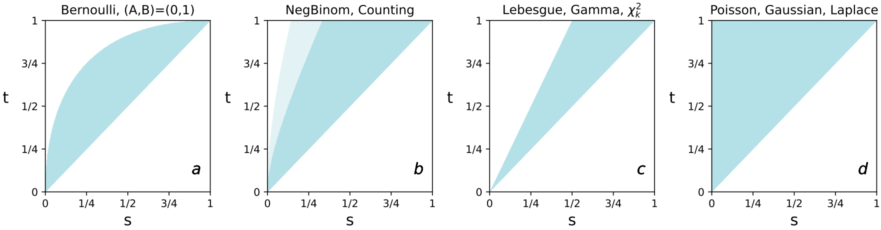


Figure: Phase diagrams for tame margins for various base measures μ. The upper contours are given by $(s+t)^2 < 2s$, $t \leq 1 + \sqrt{1 + rs^{-1}}$ ($r = 5$ for NegBinom and $r = 1$ for Counting), $t \leq s/2$, and $t = \infty$ from left to right.

Theorem (L-M '24+)

Suppose $-\infty < A \leq B < \infty$. *Then each* $(s, t) \in (A, B)^2$ *with* $s \leq t$ *belongs to* $\Omega(\mu)$ *if*

$$(s + t - 2A)^2 < 4(B - A)(s - A).$$

- A Matrix version of Erdös-Galai condition
- Universality: only depends on $B - A$

Geometry of tame margins

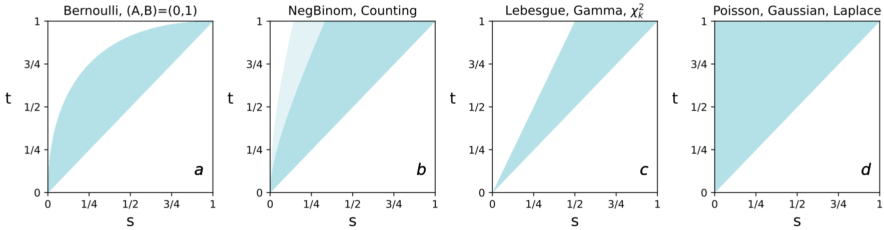


Figure: Phase diagrams for tame margins for various base measures μ. The upper contours are given by $(s+t)^2 < 2s$, $t \le 1 + \sqrt{1 + rs^{-1}}$ ($r = 5$ for NegBinom and $r = 1$ for Counting), $t \le s/2$, and $t = \infty$ from left to right.

Theorem (L-M '24+)

$(s, t) \in \Omega(\delta)$ *if and only if* $t/s < \lambda_c$ *where*

$$\lambda_c := \begin{cases} 1 + \sqrt{1 + rs^{-1}} & \text{if } \mu = r\text{-fold convolution of the counting measure on } \mathbb{Z}_{\ge 0}, \\ 2 & \text{if } \mu = \text{Gamma, Lebesgue}(\mathbb{R}_{\ge 0}), \text{ or } \chi^2_k \\ \infty & \text{if } \mu = \text{Poisson, Gaussian, Laplace} \end{cases}$$

Open problems

- (*Summary*)

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

- $X_{11} \overset{d}{=} \ \mu_{\boldsymbol{\alpha}(1)+\boldsymbol{\beta}(1)}$
- $\mathbb{E}[X] \approx \ Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = \mathbb{E}[Y]$
- $X - \mathbb{E}[Y] =$ small w.h.p. in cut norm
- $X \overset{d}{\approx} \ Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$

Open problems

- (*Summary*)

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

  - $X_{11} \overset{d}{=} \mu_{\boldsymbol{\alpha}(1)+\boldsymbol{\beta}(1)}$
  - $\mathbb{E}[X] \approx Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = \mathbb{E}[Y]$
  - $X - \mathbb{E}[Y] =$ small w.h.p. in cut norm
  - $X \overset{d}{\approx} Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$
- Empirical distribution of eigenvalues of $X - \mathbb{E}[Y]$? (circular for constant margin with $\mu = \text{Leb}(\mathbb{R}_{\geq 0})$ by Nguyen '14 [16])

Open problems

- (*Summary*)

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

  - $X_{11} \overset{d}{=} \mu_{\boldsymbol{\alpha}(1)+\boldsymbol{\beta}(1)}$
  - $\mathbb{E}[X] \approx Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = \mathbb{E}[Y]$
  - $X - \mathbb{E}[Y] =$ small w.h.p. in cut norm
  - $X \overset{d}{\approx} Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$

- Empirical distribution of eigenvalues of $X - \mathbb{E}[Y]$? (circular for constant margin with $\mu = \text{Leb}(\mathbb{R}_{\geq 0})$ by Nguyen '14 [16])
- Eigenvalue distribution with additional symmetry $X^\top = X$? (Semi-circle for constant margin, generalized Wigner matrices, quadratic vector eq. etc. ; Discussing with Hongchang Ji)

- (*Summary*)

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

  - $X_{11} \overset{d}{=} \mu_{\boldsymbol{\alpha}(1)+\boldsymbol{\beta}(1)}$
  - $\mathbb{E}[X] \approx Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = \mathbb{E}[Y]$
  - $X - \mathbb{E}[Y] =$ small w.h.p. in cut norm
  - $X \overset{d}{\approx} Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$

- Empirical distribution of eigenvalues of $X - \mathbb{E}[Y]$? (circular for constant margin with $\mu = \text{Leb}(\mathbb{R}_{\geq 0})$ by Nguyen '14 [16])
- Eigenvalue distribution with additional symmetry $X^\top = X$? (Semi-circle for constant margin, generalized Wigner matrices, quadratic vector eq. etc. ; Discussing with Hongchang Ji)
- Extend the theory for more general base measure for RM ensemble than i.i.d.? (with $\mu = \text{Poisson}(1)$, connection to entropic optimal transport; Ongoing work with William Powell)

Open problems

- (*Summary*)

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

  - $X_{11} \overset{d}{=} \mu_{\boldsymbol{\alpha}(1)+\boldsymbol{\beta}(1)}$
  - $\mathbb{E}[X] \approx Z^{\mathbf{r},\mathbf{c}} = \psi'(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = \mathbb{E}[Y]$
  - $X - \mathbb{E}[Y] =$ small w.h.p. in cut norm
  - $X \overset{d}{\approx} Y \sim \boldsymbol{\mu}_{\boldsymbol{\alpha}\oplus\boldsymbol{\beta}}$
- Empirical distribution of eigenvalues of $X - \mathbb{E}[Y]$? (circular for constant margin with $\mu = \text{Leb}(\mathbb{R}_{\geq 0})$ by Nguyen '14 [16])
- Eigenvalue distribution with additional symmetry $X^{\top} = X$? (Semi-circle for constant margin, generalized Wigner matrices, quadratic vector eq. etc. ; Discussing with Hongchang Ji)
- Extend the theory for more general base measure for RM ensemble than i.i.d.? (with $\mu =$ Poisson(1), connection to entropic optimal transport; Ongoing work with William Powell)
- DLP? (For random graphs with given degree sequence, LDP is done by Dhara and Sen '22 [6])
  - Ongoing work with Sumit Mukherjee

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. "Random graphs with a given degree sequence". In: *The Annals of Applied Probability* 21.4 (2011), pp. 1400–1435.

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

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

[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] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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

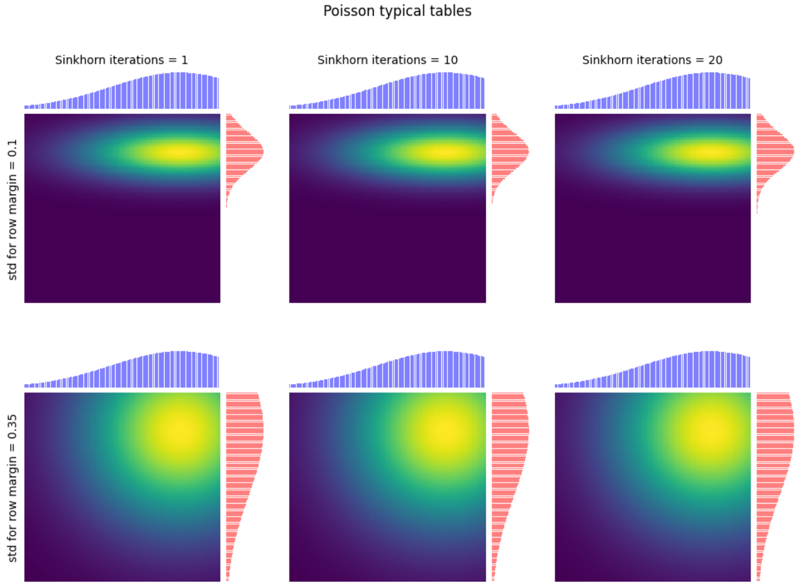
$$\begin{cases} \text{For } 1 \le i \le m, \; \boldsymbol{\alpha}_k(i) \leftarrow \text{unique } \alpha \text{ s.t. } \mathbf{r}(i) = \sum_{j=1}^n \psi'(\alpha + \boldsymbol{\beta}_{k-1}(j)), \\ \text{For } 1 \le j \le n, \; \boldsymbol{\beta}_k(j) \leftarrow \text{unique } \beta \text{ s.t. } \mathbf{c}(j)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

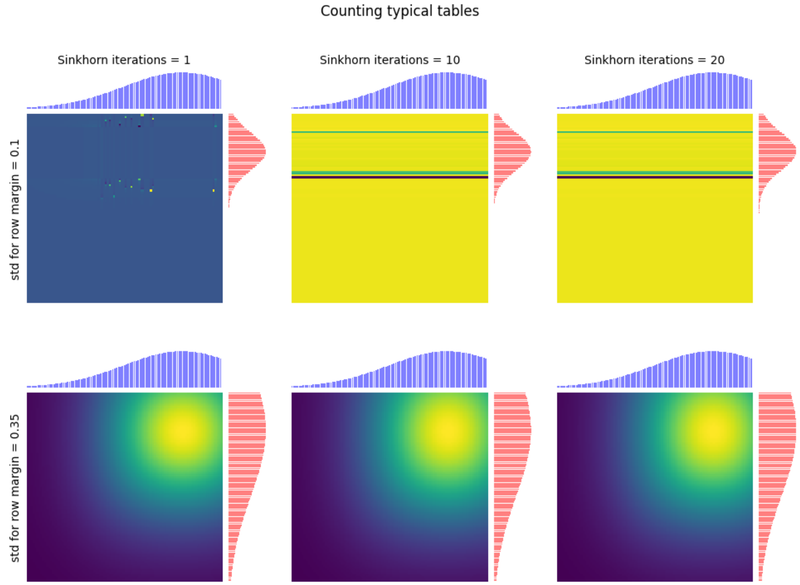
  - For $\mu = \text{Poisson}(1)$ (Schrödinger bridge), $\psi'(x) = e^x$, so

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

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 $\iff (\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ 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 $\iff$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ 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 $\Longleftrightarrow$ $(\boldsymbol{\alpha}^* + \lambda, \boldsymbol{\beta}^* - \lambda)$ 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

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
    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

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

- 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
    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*

Outline

- 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. $\mathcal{R}$ [9, 17]

- 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. $\mathcal{R}$ [9, 17]

- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [18] 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. $\mathcal{R}$ [9, 17]
- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [18] 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. $\mathcal{R}$ [9, 17]

- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [18] 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. $\mathcal{R}$ [9, 17]
- $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 : \mathbb{R} \to \mathbb{R}$, the **Schrödinger potentials** [18] 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 $\mu = \text{Poisson}(1)$!
  - $x_{ij}\log x_{ij} = D(\mu_{\phi(x_{ij})}\,\|\,\mu) + x_{ij} - 1$