Sinkhorn's Algorithm and Random Matrices: Duality, Stability, and Lagrangian Costs

Hanbaek Lyu

Department of Mathematics

University of Wisconsin – Madison

Partially supported by NSF CAREER DMS-2541878

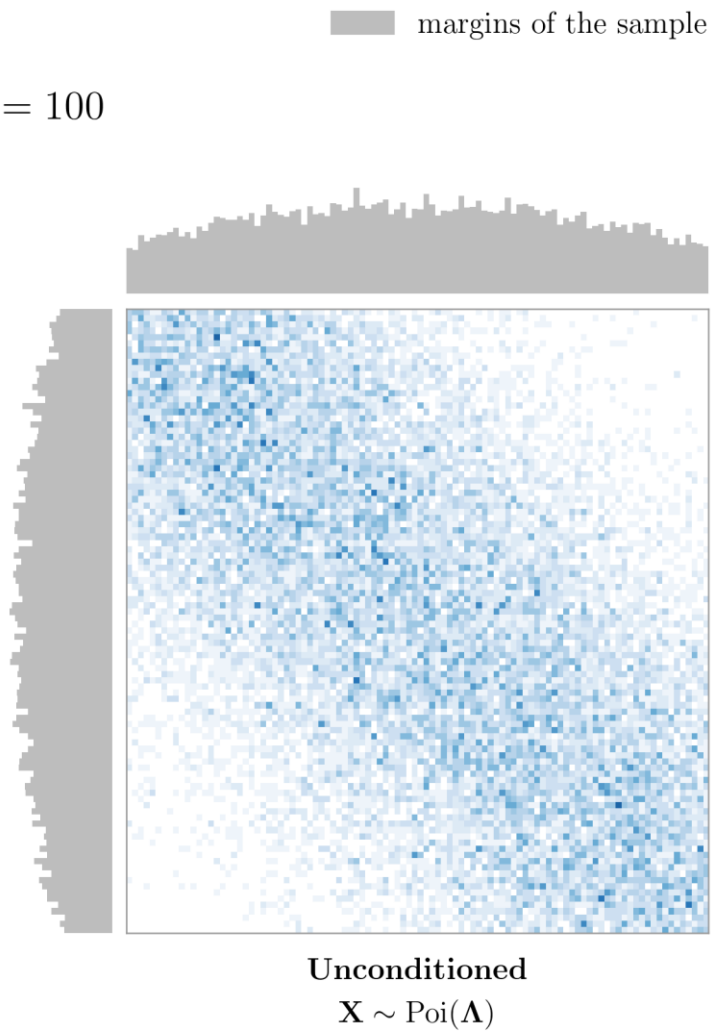
Midwest Optimization Meeting 2026

South Dakota State University, October 30–31, 2026

Introduction

Random Matrices Conditioned on Margin

Poisson Contingency Tables



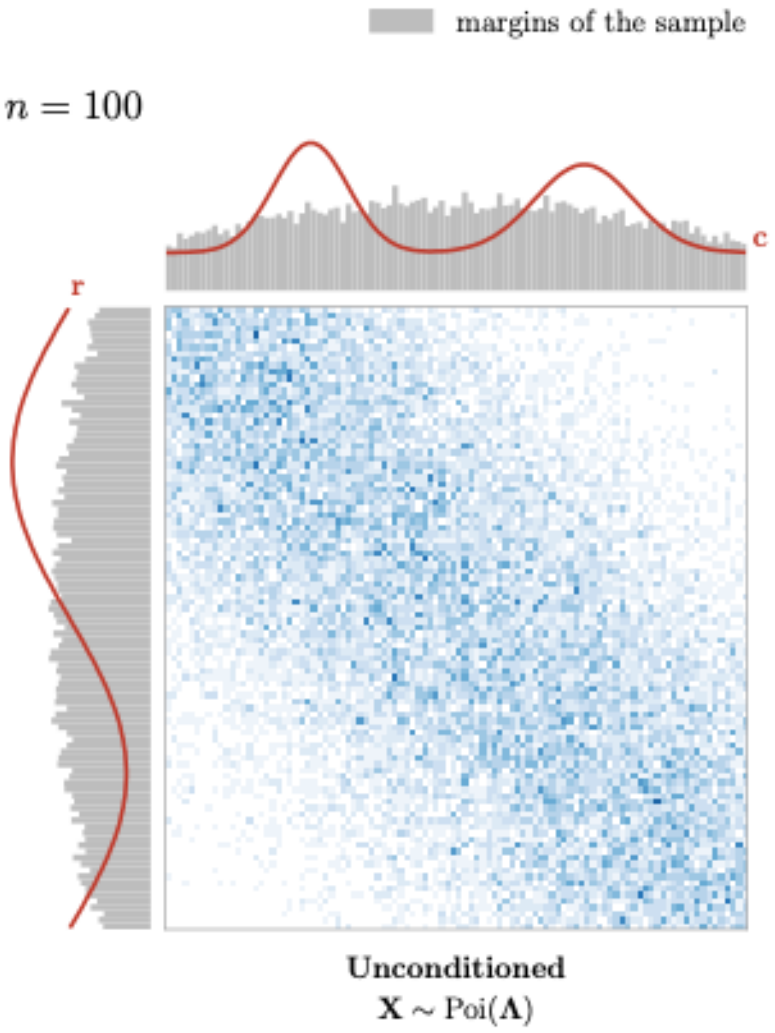
- $\mathbf{\Lambda} = (\mathbf{\Lambda_{ij}})_{n \times n}$

  Nonnegative prior mean matrix

- $\boldsymbol{X} \sim \mathrm{Poisson}(\mathbf{\Lambda})$

  $(\boldsymbol{X}_{ij} \sim \mathrm{Poisson}(\mathbf{\Lambda_{ij}})$, independent)

Poisson Contingency Tables



- $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda_{ij}})_{n\times n}$

  Nonnegative prior mean matrix

- $\boldsymbol{X} \sim \mathrm{Poisson}(\boldsymbol{\Lambda})$

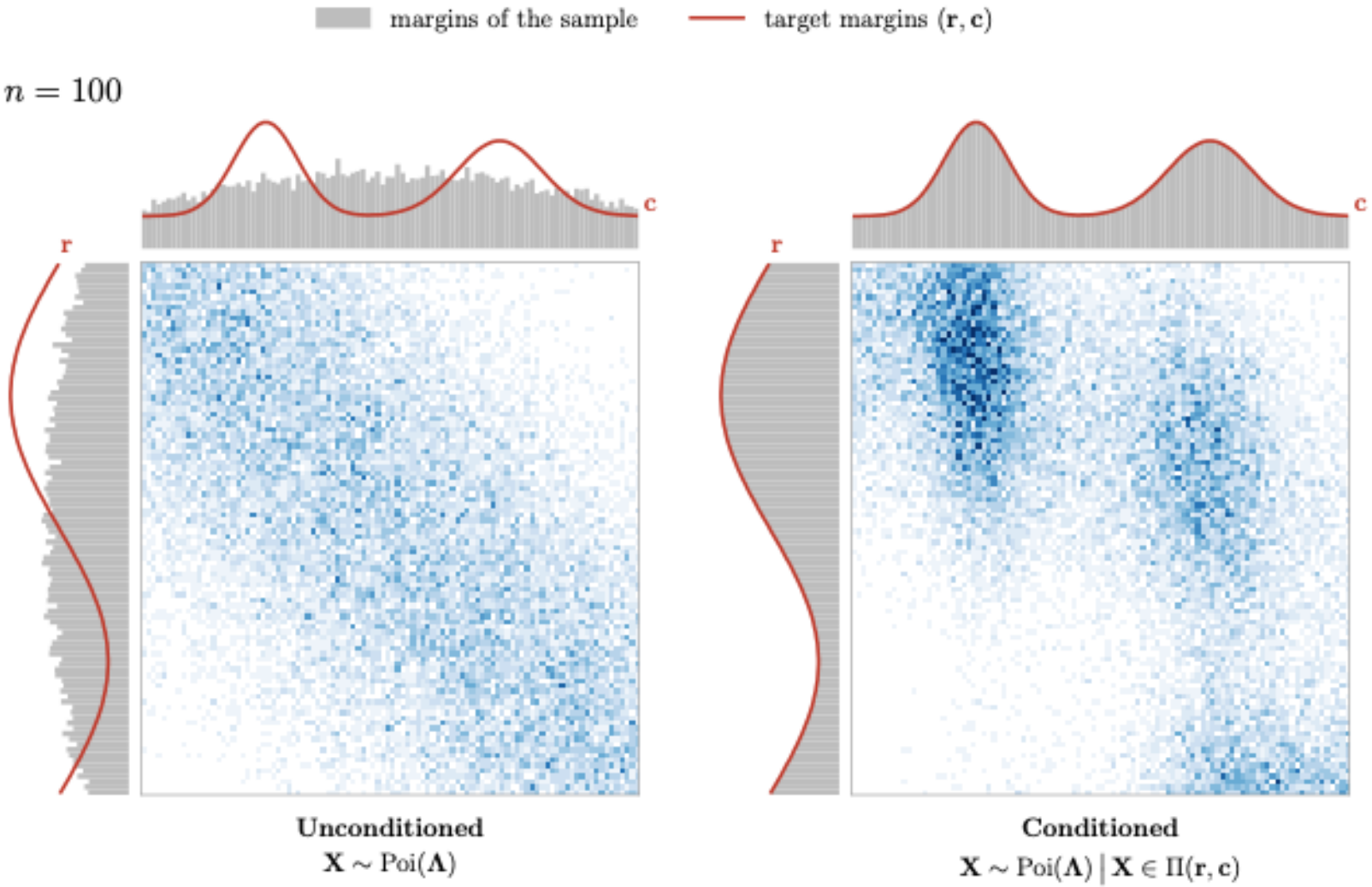
  ($\boldsymbol{X_{ij}} \sim \mathrm{Poisson}(\boldsymbol{\Lambda_{ij}})$, independent)

- $\boldsymbol{r} = (\boldsymbol{r_1}, \dots, \boldsymbol{r_n})$
  $\boldsymbol{c} = (\boldsymbol{c_1}, \dots, \boldsymbol{c_n})$ Target margin

  $\Pi(\boldsymbol{r}, \boldsymbol{c})$ = Transportation polytope

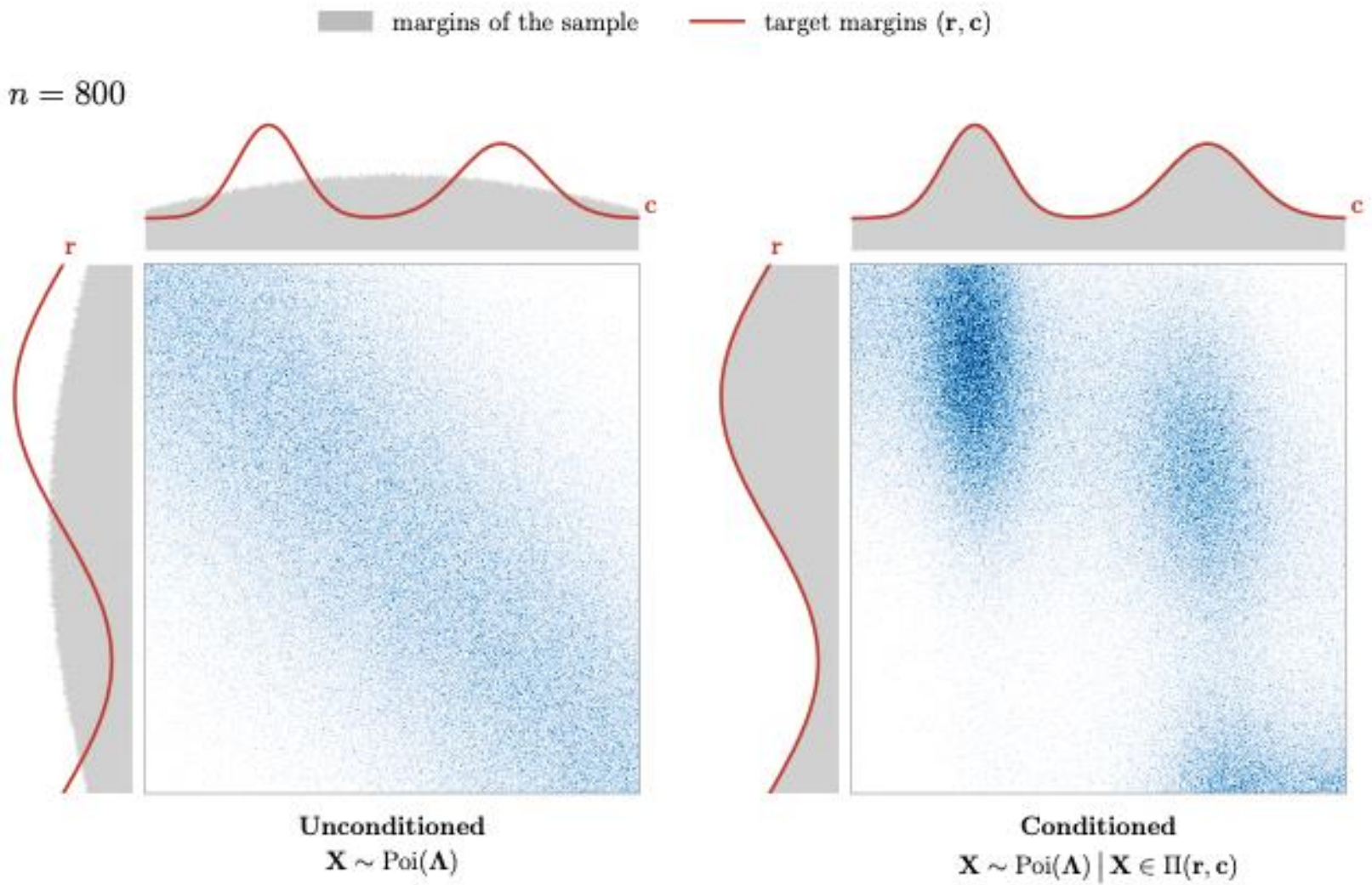
  (Set of all nonnegative matrices with margin $(\boldsymbol{r}, \boldsymbol{c})$)

Poisson Contingency Tables

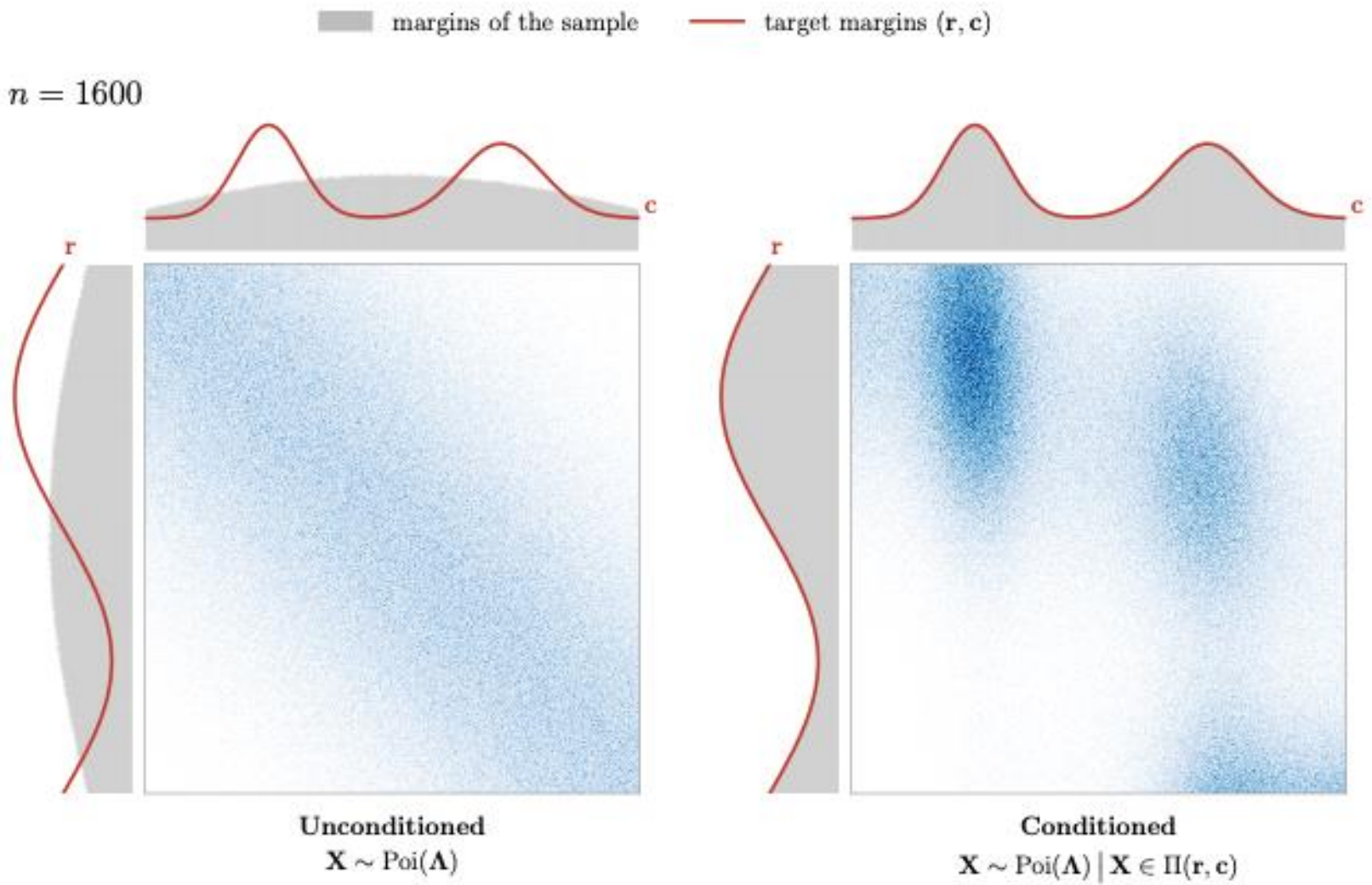


(Sampled through uniform random permutation and block counting)

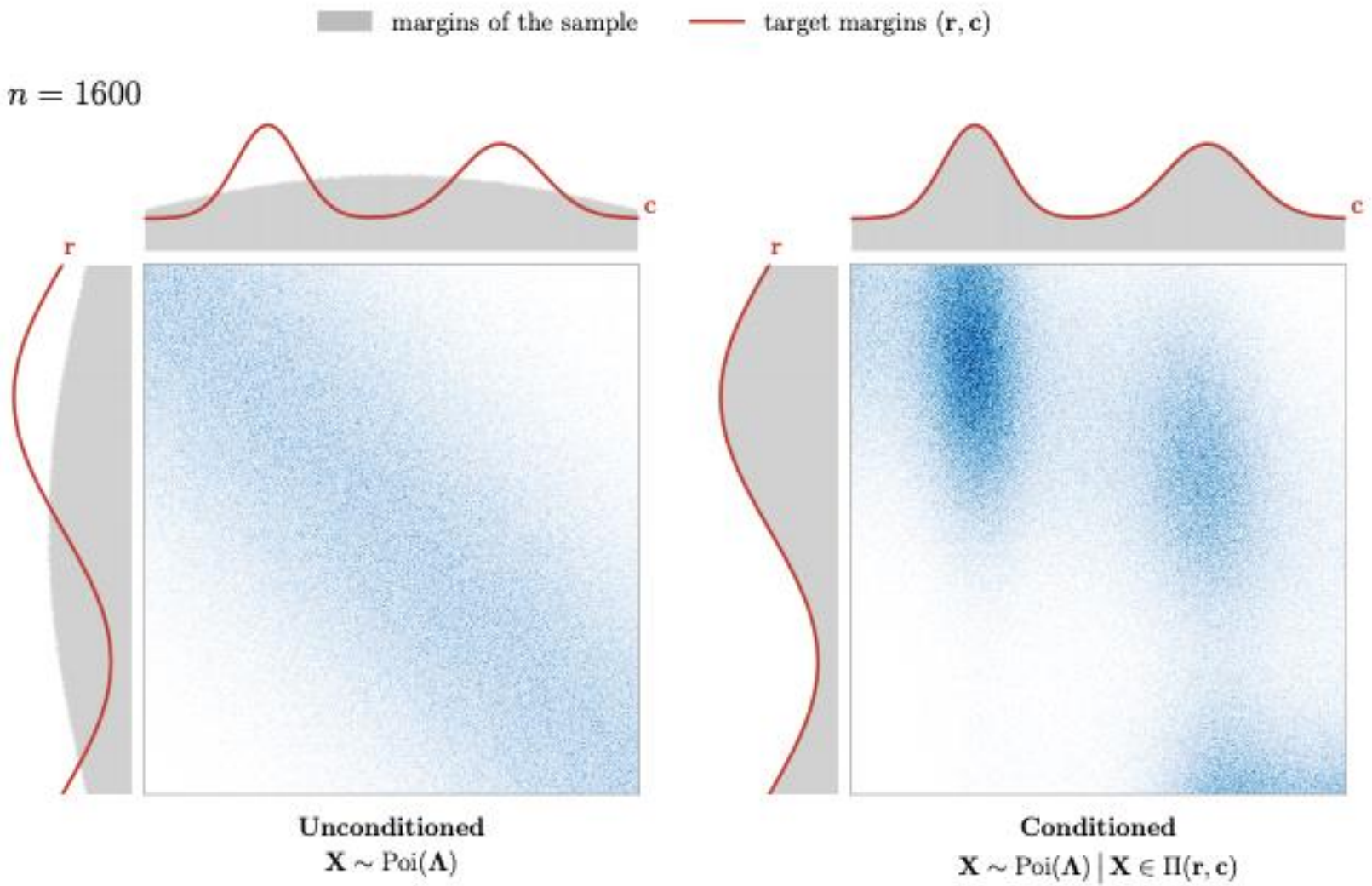
Poisson Contingency Tables



Poisson Contingency Tables

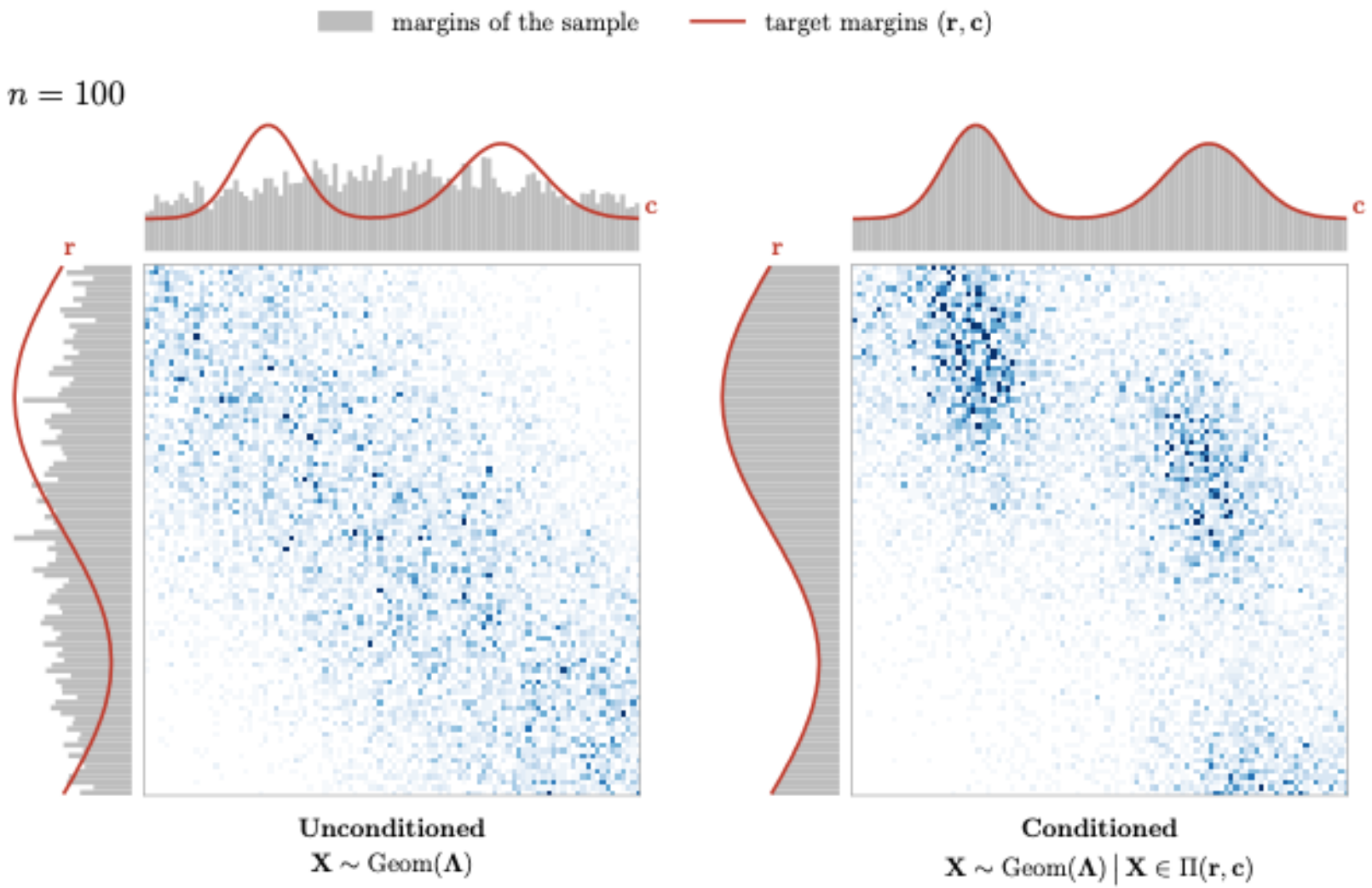


Poisson Contingency Tables

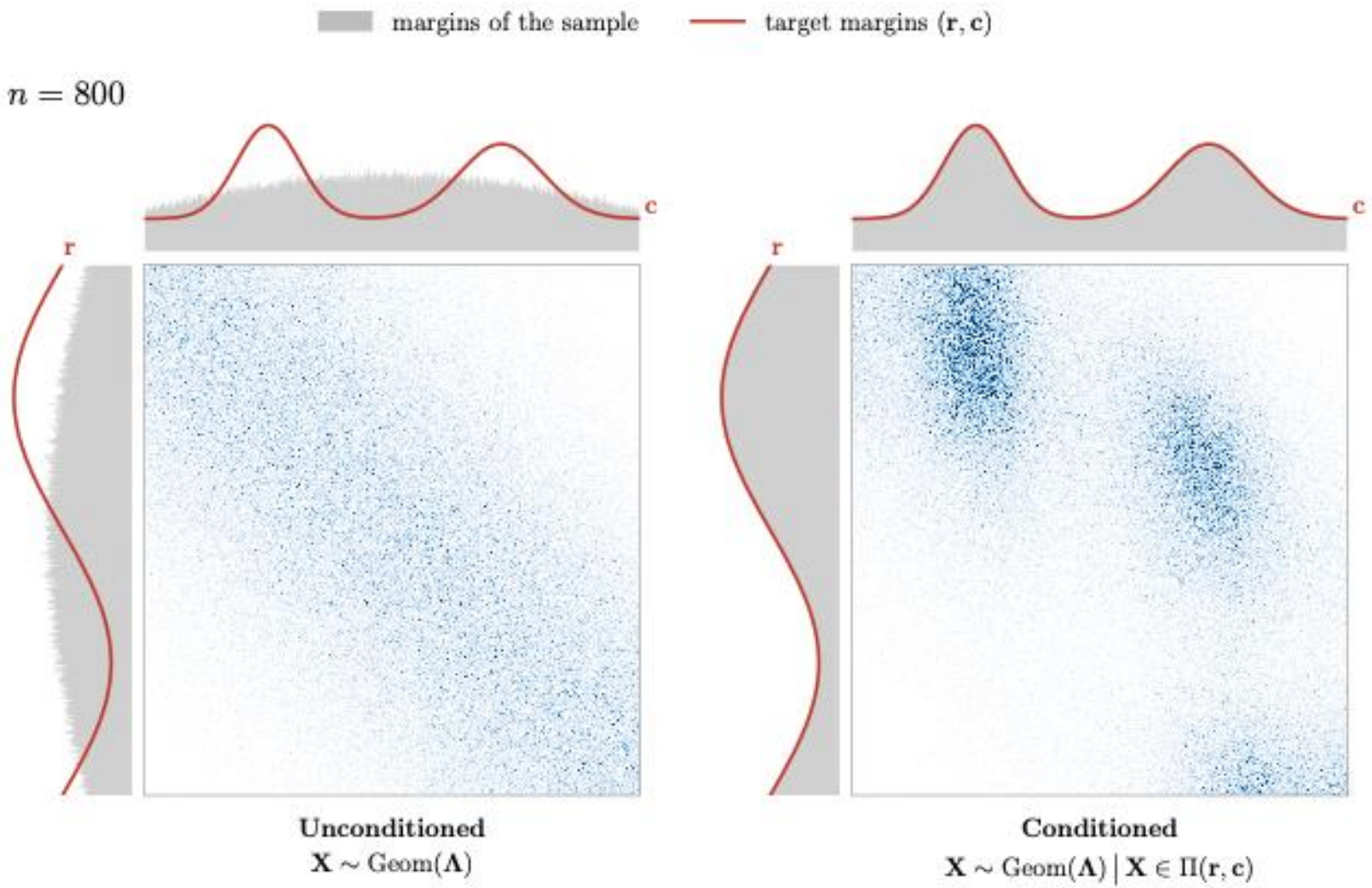


Converges to some deterministic limit as $n \to \infty$? YES, Entropic OT

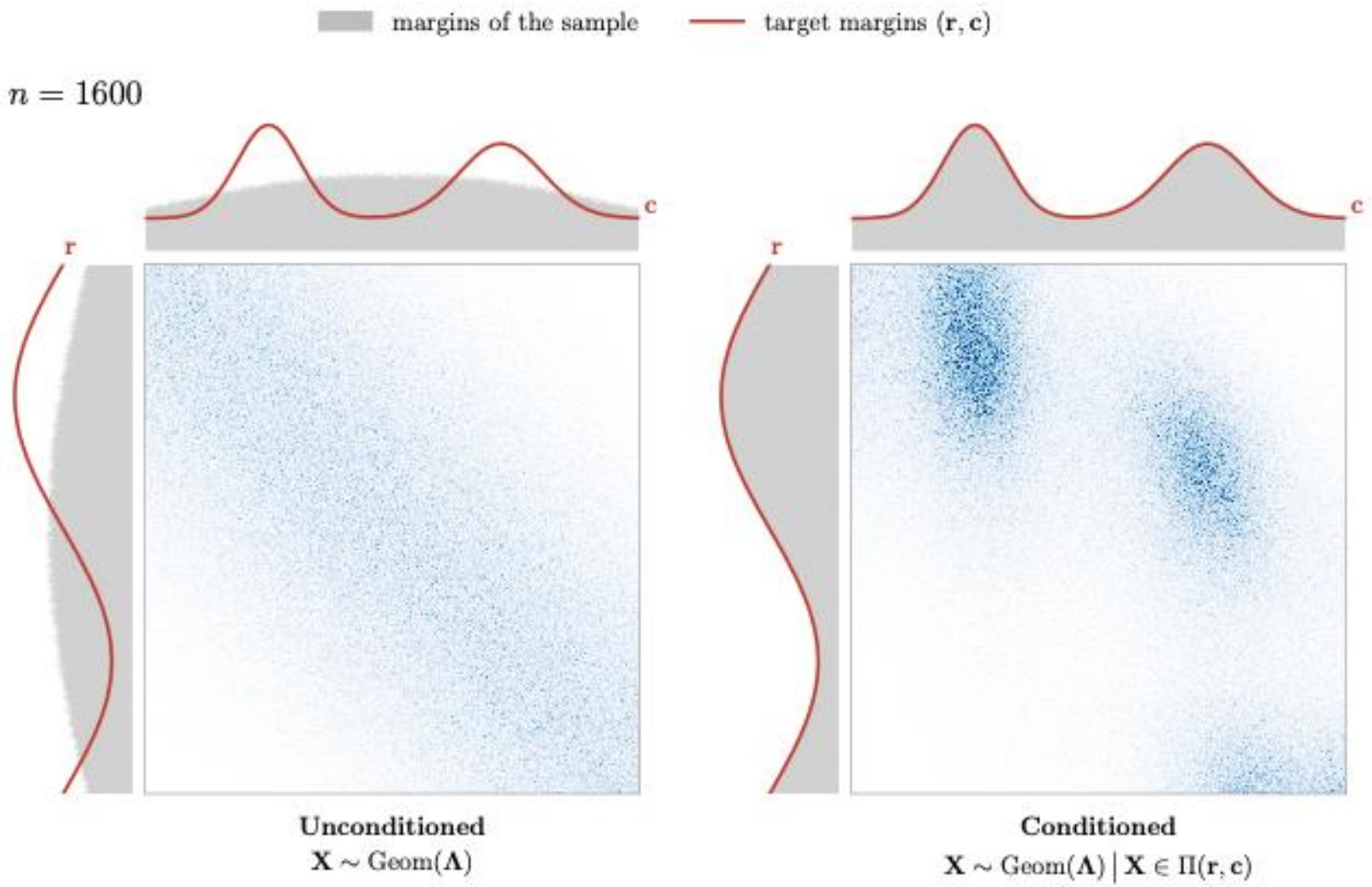
Geometric Contingency Tables



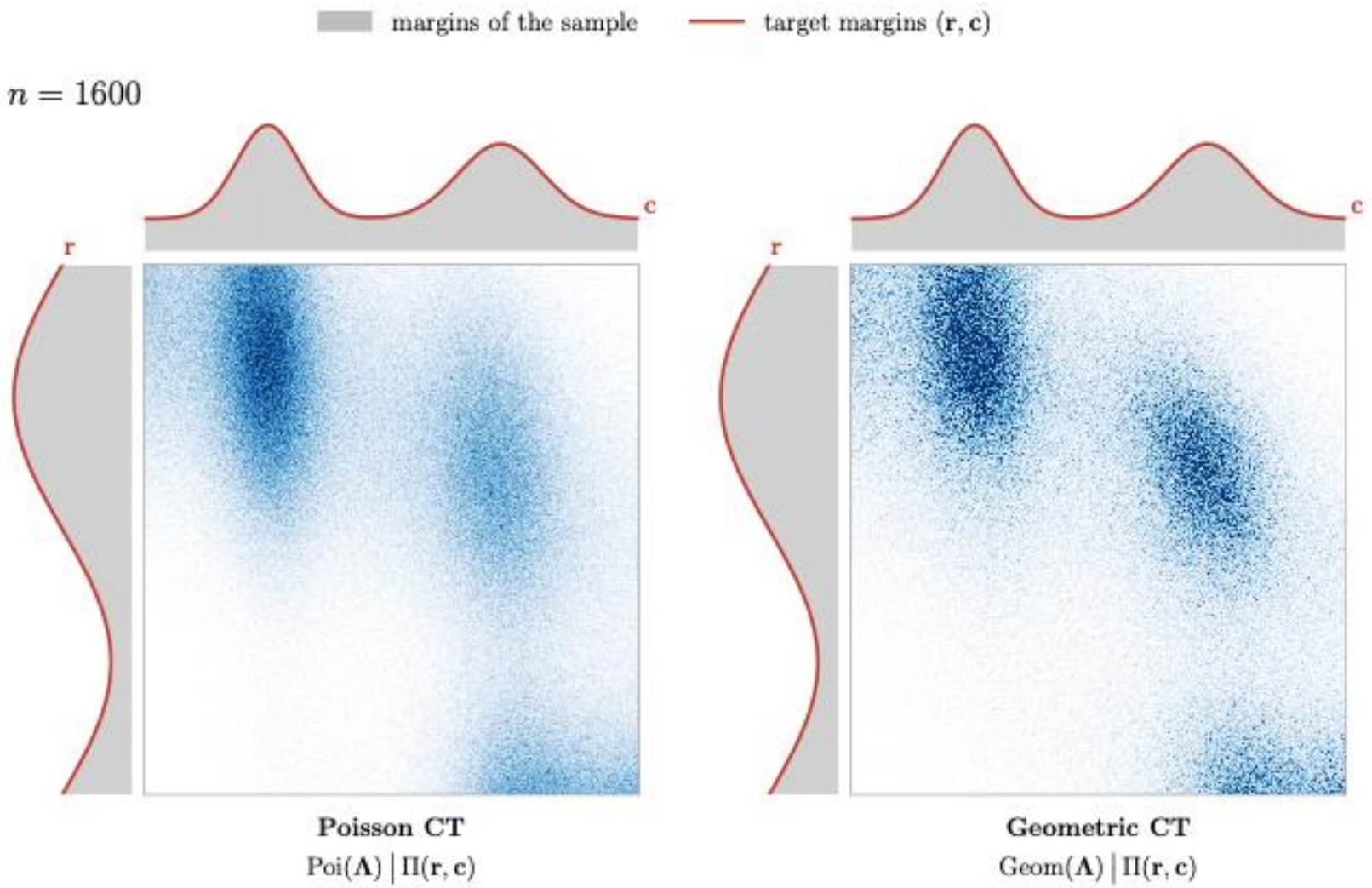
Geometric Contingency Tables



Geometric Contingency Tables



Poisson/Geometric Contingency Tables



Converge to different deterministic limits as $n \to \infty$?
YES, Static Schrödinger Bridges with KL/JS – divergences

Matrix Scaling and Sinkhorn's Algorithm

- Matrix Scaling Problem:

  Given a nonnegative matrix $\mathbf{A}$ and target row and column sums ($\mathbf{r}$, $\mathbf{c}$),

  find diagonal matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ s.t. $\mathbf{D_1 A D_2}$ satisfies margin ($\mathbf{r}$, $\mathbf{c}$)

- Matrix Scaling Problem:

  Given a nonnegative matrix **A** and target row and column sums (**r**, **c**),

  find diagonal matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ s.t. $\mathbf{D_1 A D_2}$ satisfies margin (**r**, **c**)

  - Sinkhorn's matrix scaling algorithm (1964)
    - Normalize the rows of **A** to match the target row sum **r;** Obtain **A'**
    - Normalize the columns of **A'** to match the target column sum **c;** Obtain **A"**
    - Repeat

- Matrix Scaling Problem:

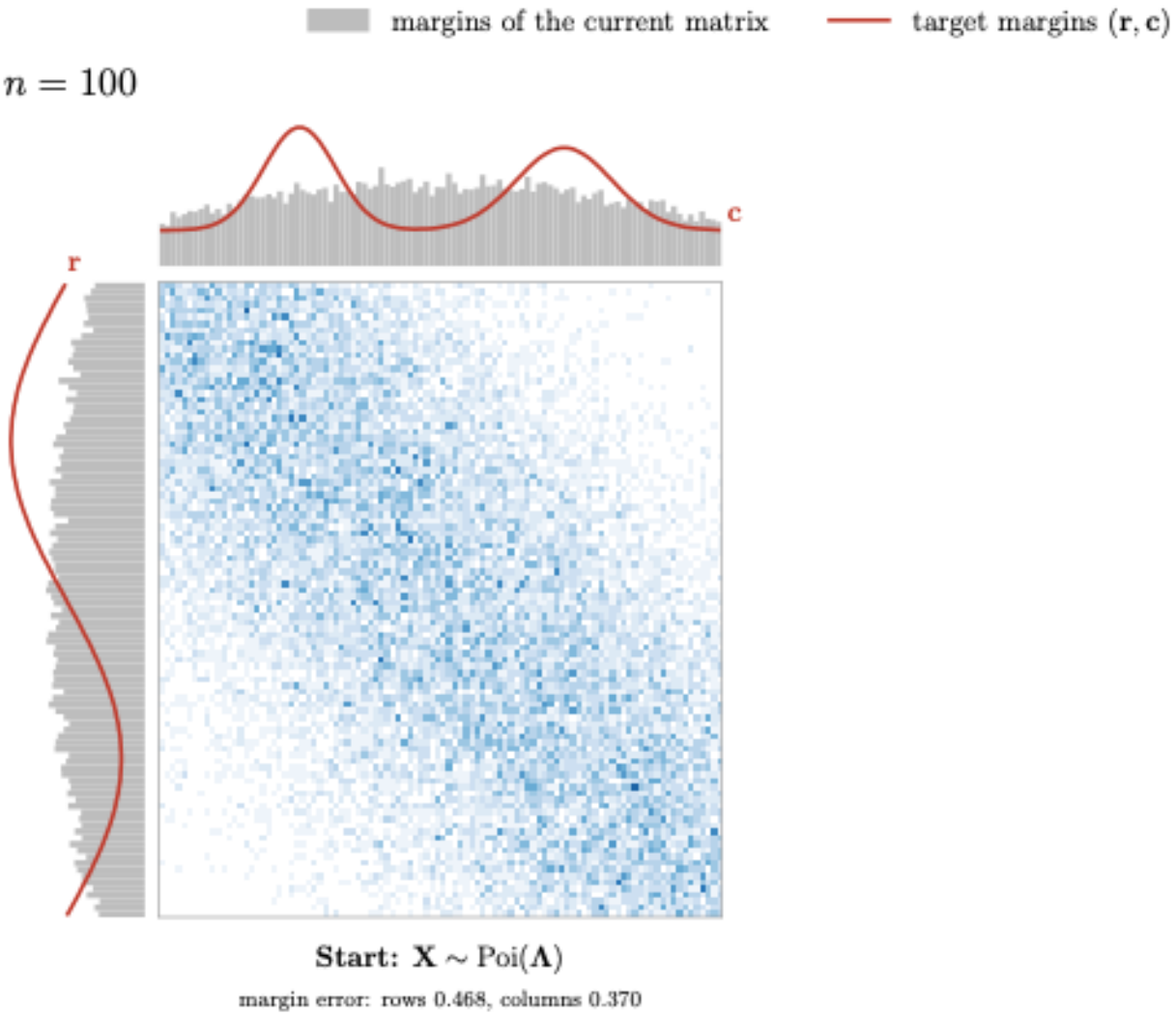
  Given a nonnegative matrix **A** and target row and column sums (**r**, **c**), find diagonal matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ s.t. $\mathbf{D_1 A D_2}$ satisfies margin (**r**, **c**)

  - Sinkhorn's matrix scaling algorithm (1964)
    - Normalize the rows of **A** to match the target row sum **r;** Obtain **A'**
    - Normalize the columns of **A'** to match the target column sum **c;** Obtain **A"**
    - Repeat
  - Sinkhorn's algorithm is known to solve the following relative entropy minimization problem:

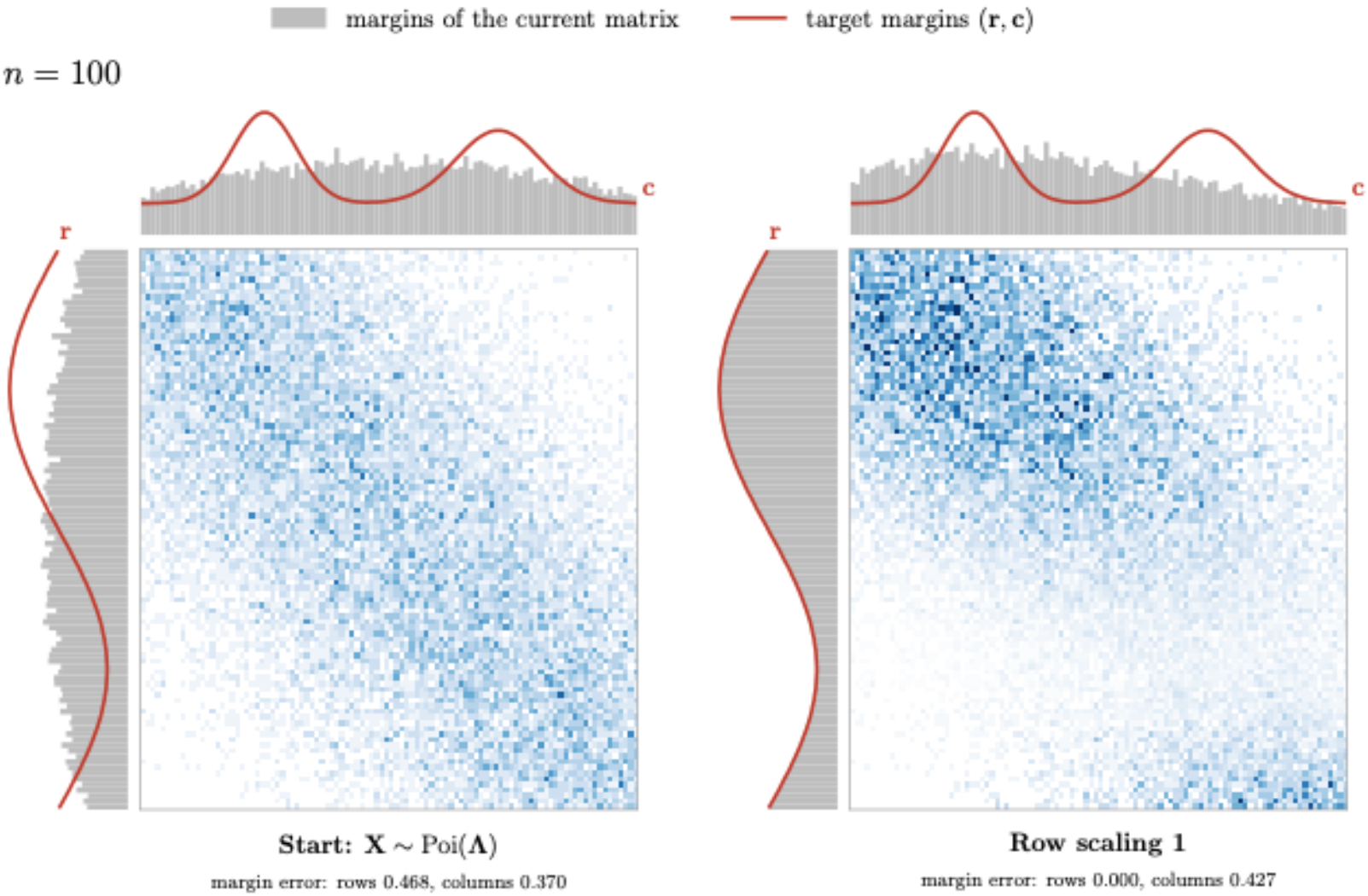
$$\underset{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})}{\arg\min}\left[D_{KL}(\mathbf{Z}\|\mathbf{A})=\sum_{ij}\mathbf{Z}_{ij}\log\left(\mathbf{Z}_{ij}/\mathbf{A}_{ij}\right)\right]$$

Transportation polytope with margin (**r**, **c**)

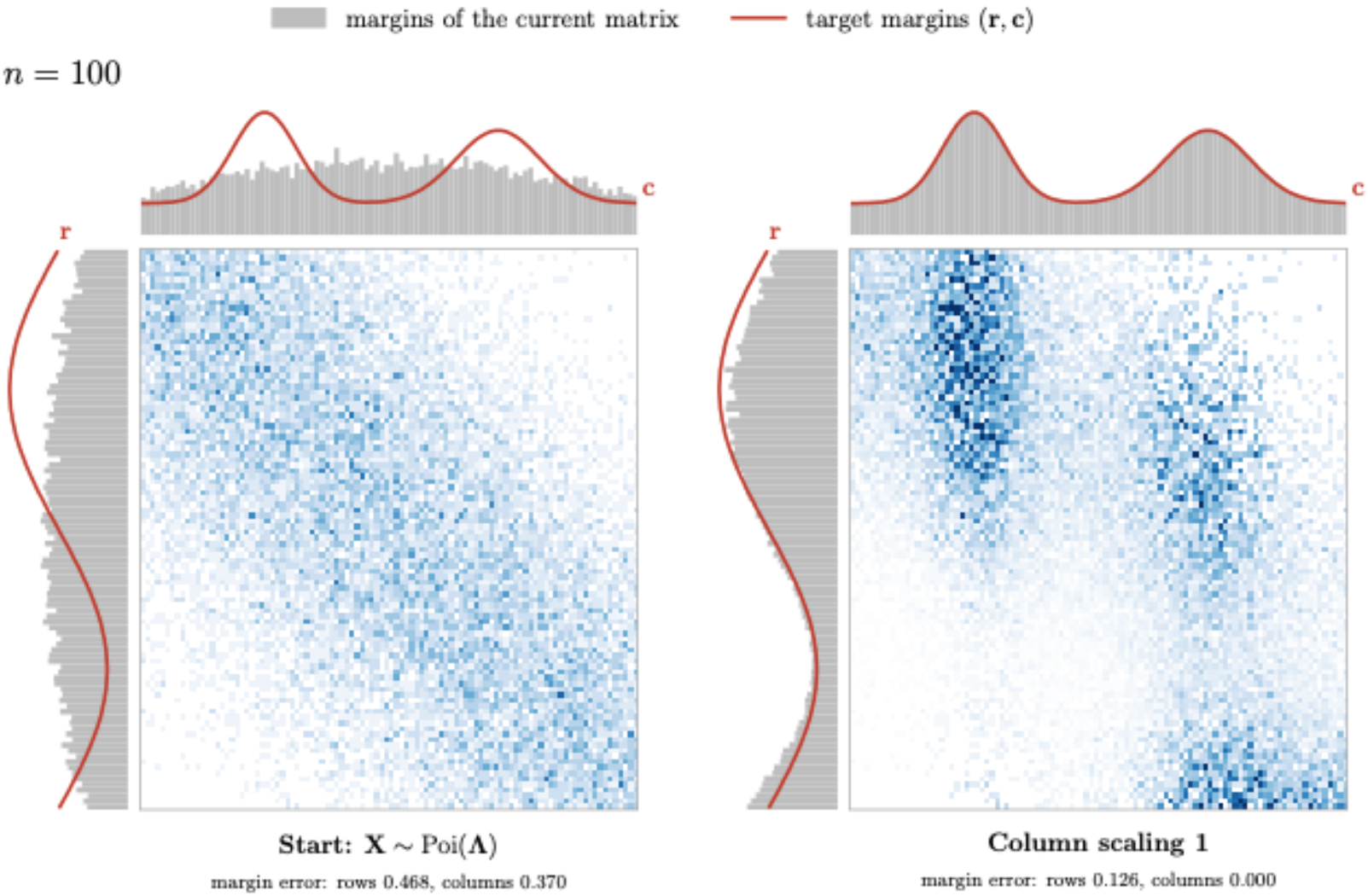
Sinkhorn-rescaled Poisson Matrices



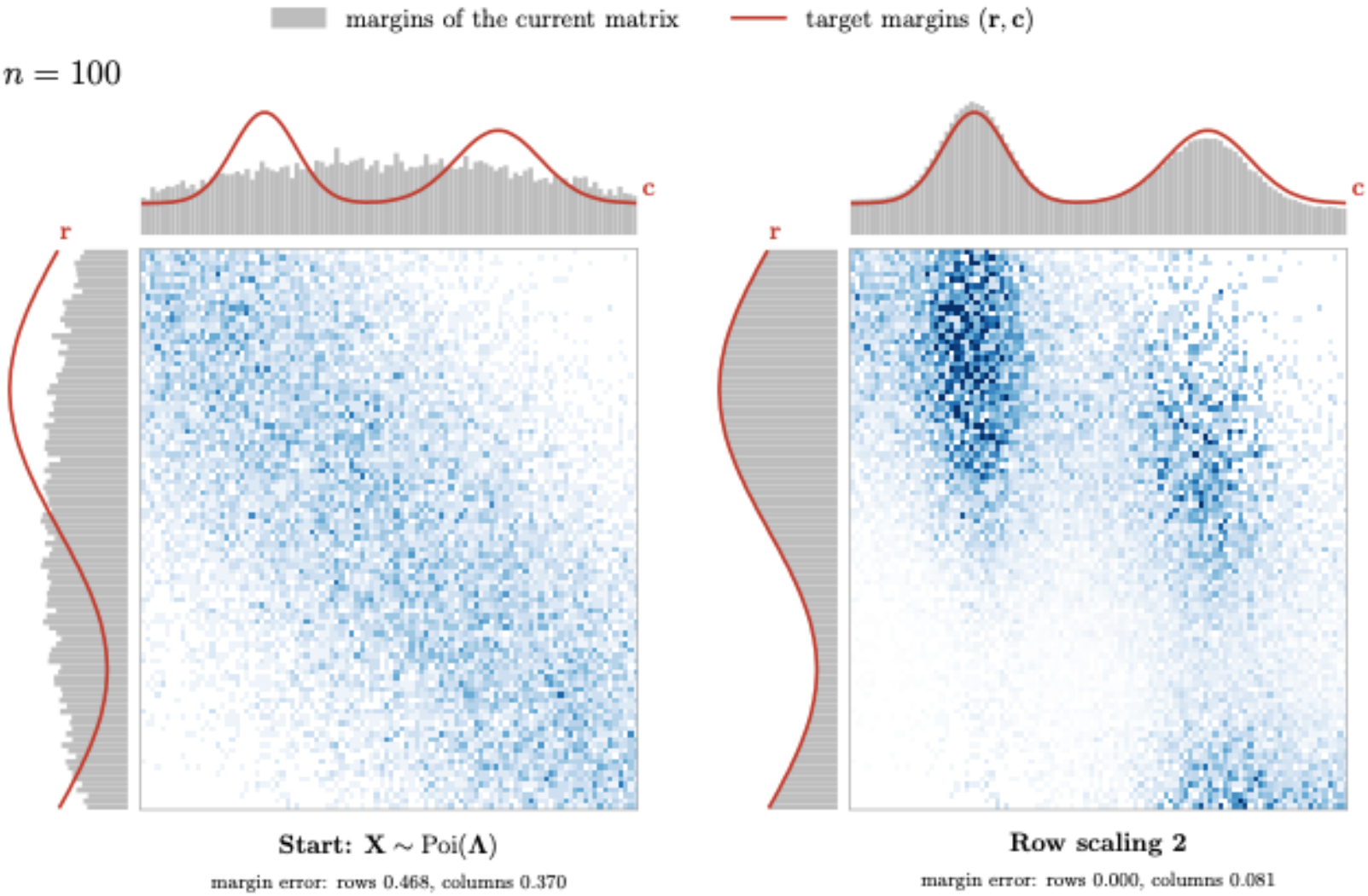
Sinkhorn-rescaled Poisson Matrices



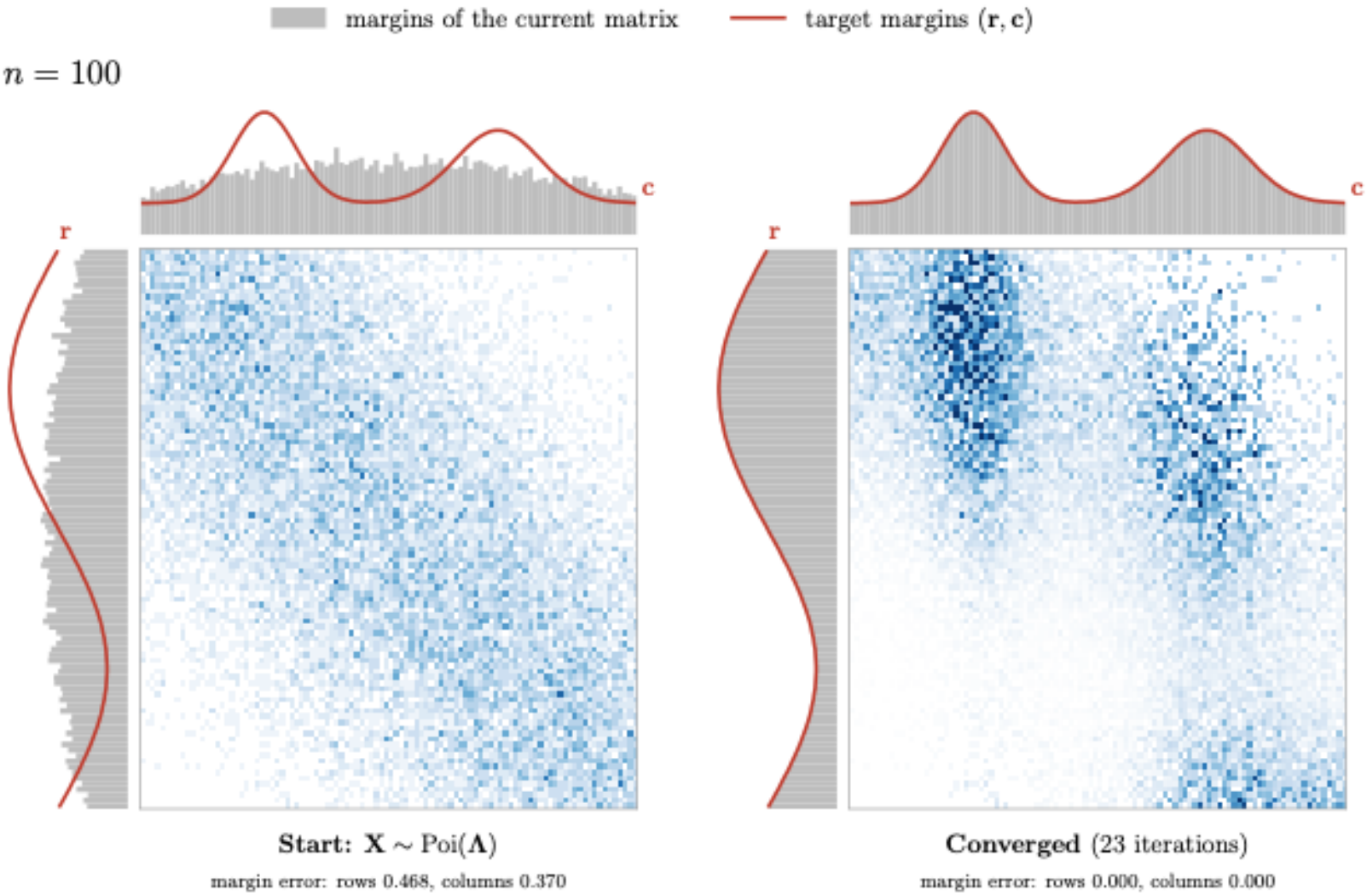
Sinkhorn-rescaled Poisson Matrices



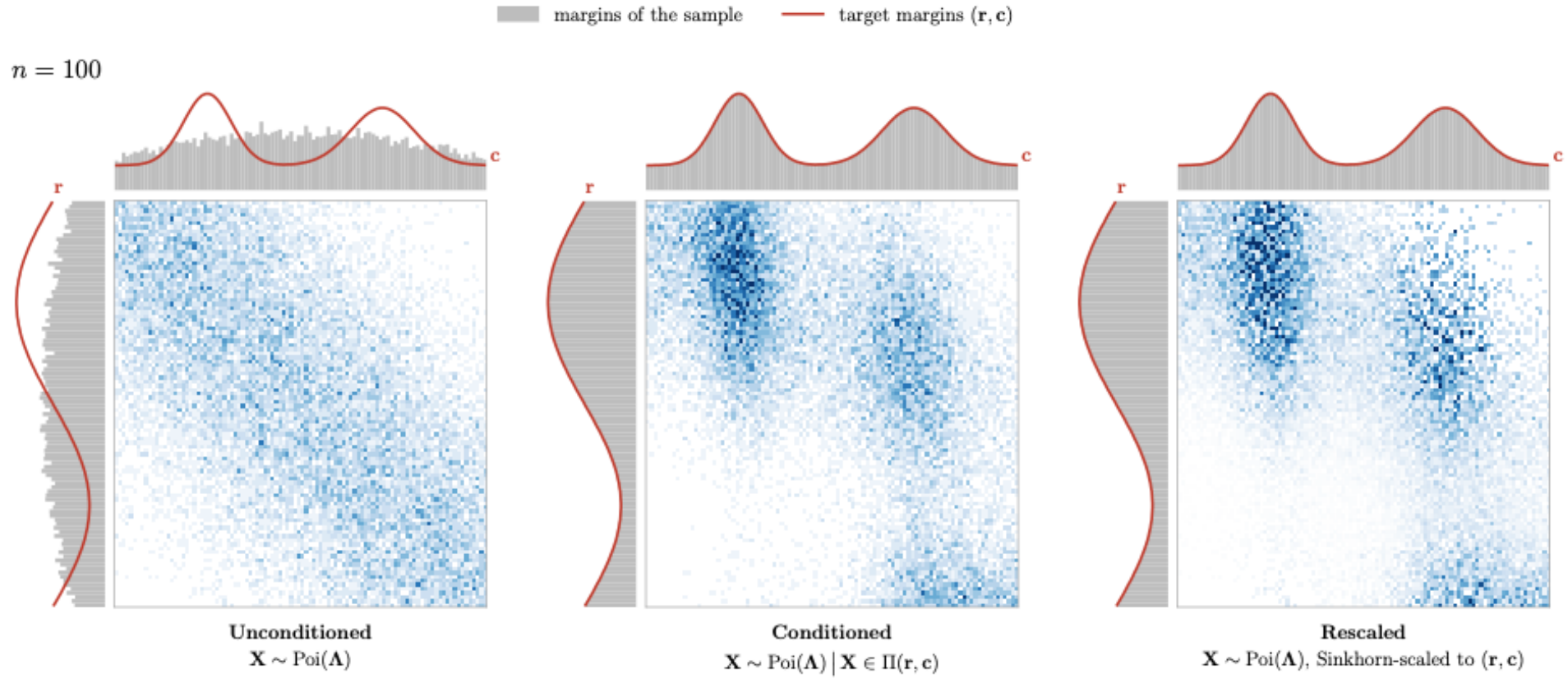
Sinkhorn-rescaled Poisson Matrices



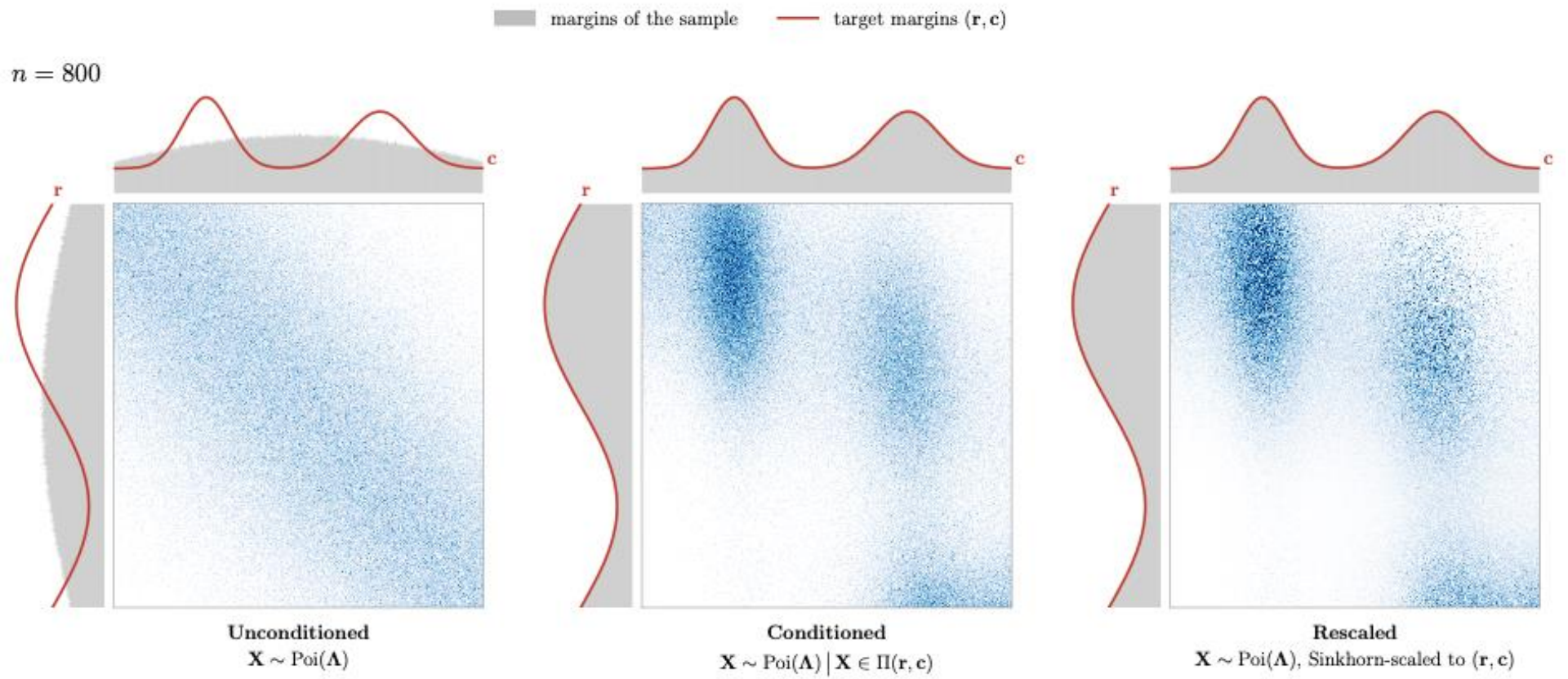
Sinkhorn-rescaled Poisson Matrices



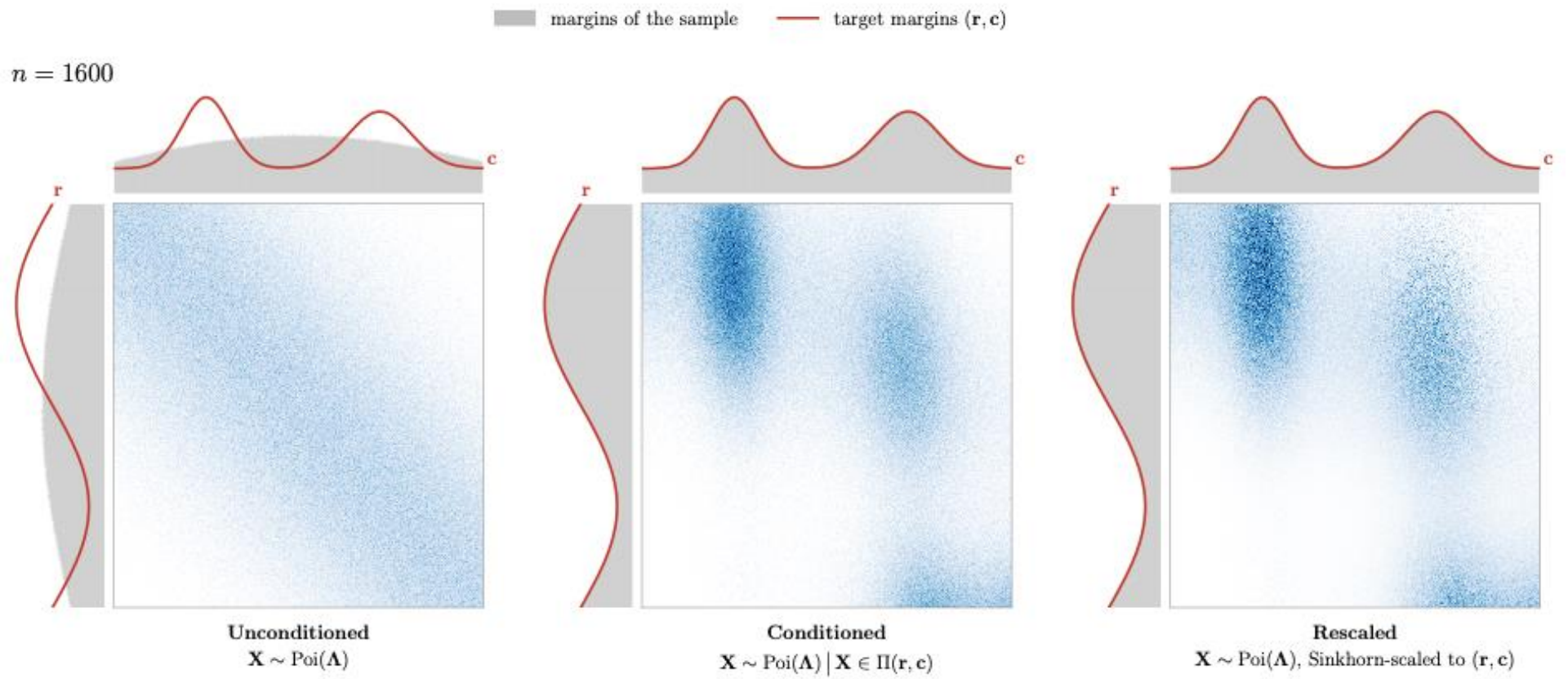
Poisson Matrices conditioned vs. rescaled



Poisson Matrices conditioned vs. rescaled



Poisson Matrices conditioned vs. rescaled



Converges to the **same** deterministic limit as $n \to \infty$?

YES!

Duality
Optimal Transport
Schrödinger Bridges

- Sinkhorn's matrix scaling algorithm (1964)
  - Normalize the rows of **A** to match the target row sum **r;** Obtain **A'**
  - Normalize the columns of **A'** to match the target column sum **c;** Obtain **A"**
  - Repeat

- Sinkhorn's algorithm is known to solve the following relative entropy minimization problem:

$$\underset{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})}{\arg\min}\left[D_{KL}(\mathbf{Z}\|\mathbf{A})=\sum_{ij}\mathbf{Z}_{ij}\log\left(\mathbf{Z}_{ij}/\mathbf{A}_{ij}\right)\right]$$

Transportation polytope with margin (**r**, **c**)

- Sinkhorn's matrix scaling algorithm (1964)
  - Normalize the rows of **A** to match the target row sum **r;** Obtain **A'**
  - Normalize the columns of **A'** to match the target column sum **c;** Obtain **A"**
  - Repeat

- Sinkhorn's algorithm is known to solve the following relative entropy minimization problem:

$$\underset{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})}{\arg\min}\left[D_{KL}(\mathbf{Z}\|\mathbf{A})=\sum_{ij}\mathbf{Z}_{ij}\log\left(\mathbf{Z}_{ij}/\mathbf{A}_{ij}\right)\right]$$

Transportation polytope with margin (**r**, **c**)

Why? It is in fact the **alternating maximization** on the "dual"!

- Kantorovich Duality and Sinkhorn's algorithm

Primal

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z}\|\mathbf{A}) \right]$$

Rescaled Matrix

- Kantorovich Duality and Sinkhorn's algorithm

Primal **Dual**

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z}\|\mathbf{A})\right] = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[\langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \boldsymbol{\beta}, \mathbf{c}\rangle - \langle \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}), \mathbf{A}\rangle\right]$$

Rescaled Matrix

$\boldsymbol{\alpha}$ = Lagrange multipliers for the row-sum constraint

$\boldsymbol{\beta}$ = Lagrange multipliers for the column-sum constraint

(Schrödinger potentials)

- Kantorovich Duality and Sinkhorn's algorithm

Primal **Dual**

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z}\|\mathbf{A})\right] = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[\langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \boldsymbol{\beta}, \mathbf{c}\rangle - \langle \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}), \mathbf{A}\rangle\right]$$

Rescaled Matrix

$\boldsymbol{\alpha}$ = Lagrange multipliers for the row-sum constraint

$\boldsymbol{\beta}$ = Lagrange multipliers for the column-sum constraint

(Schrödinger potentials)

$$\pi^* = \mathrm{Diag}(\exp(\boldsymbol{\alpha}^*))\,\mathbf{A}\,\mathrm{Diag}(\exp(\boldsymbol{\beta}^*))$$

- Kantorovich Duality and Sinkhorn's algorithm

Primal **Dual**

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z}\|\mathbf{A})\right] = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[\langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \boldsymbol{\beta}, \mathbf{c}\rangle - \langle \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}), \mathbf{A}\rangle\right]$$

Rescaled Matrix

$\boldsymbol{\alpha}$ = Lagrange multipliers for the row-sum constraint

$\boldsymbol{\beta}$ = Lagrange multipliers for the column-sum constraint

(Schrödinger potentials)

$$\pi^* = \mathrm{Diag}(\exp(\boldsymbol{\alpha}^*))\,\mathbf{A}\,\mathrm{Diag}(\exp(\boldsymbol{\beta}^*))$$

Matrix scaling algorithm
(Iterative Proportional Fitting)

Fit the row sums;
Fit the column sums; etc.

Kantorovich Duality and Sinkhorn's algorithm

Primal **Dual**

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z}\|\mathbf{A})\right] = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[\langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \boldsymbol{\beta},\mathbf{c}\rangle - \langle \exp(\boldsymbol{\alpha}\oplus\boldsymbol{\beta}),\mathbf{A}\rangle\right]$$

Rescaled Matrix

$\boldsymbol{\alpha}$ = Lagrange multipliers for the row-sum constraint

$\boldsymbol{\beta}$ = Lagrange multipliers for the column-sum constraint

(Schrödinger potentials)

$$\pi^* = \mathrm{Diag}(\exp(\boldsymbol{\alpha}^*))\,\mathbf{A}\,\mathrm{Diag}(\exp(\boldsymbol{\beta}^*))$$

Alternating maximization

Matrix scaling algorithm
(Iterative Proportional Fitting)

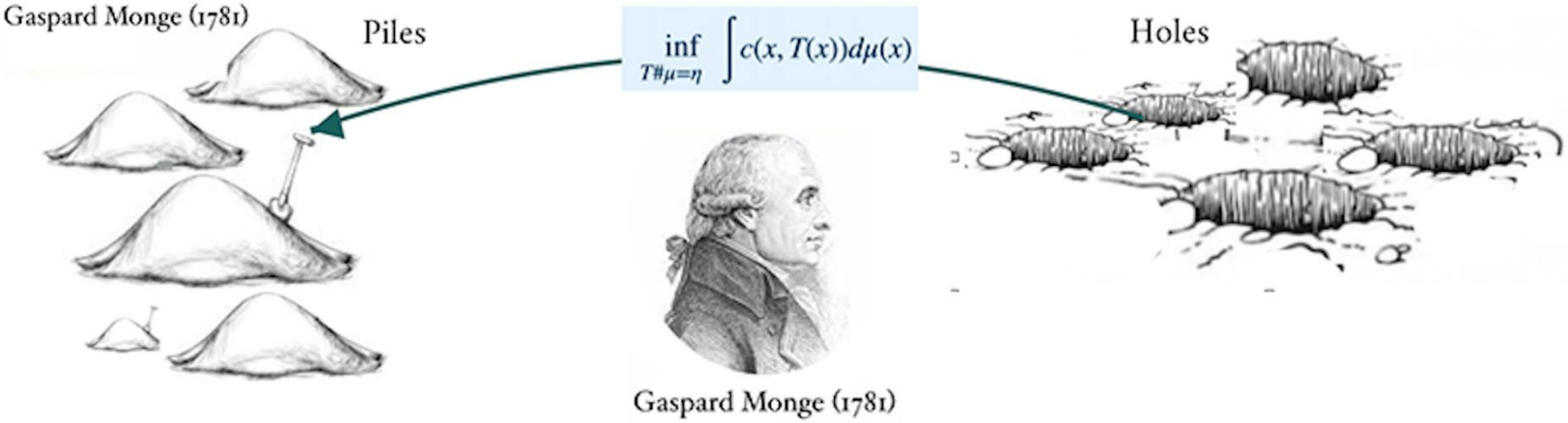
Fit the row sums;
Fit the column sums; etc.

Sinkhorn Algorithm

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

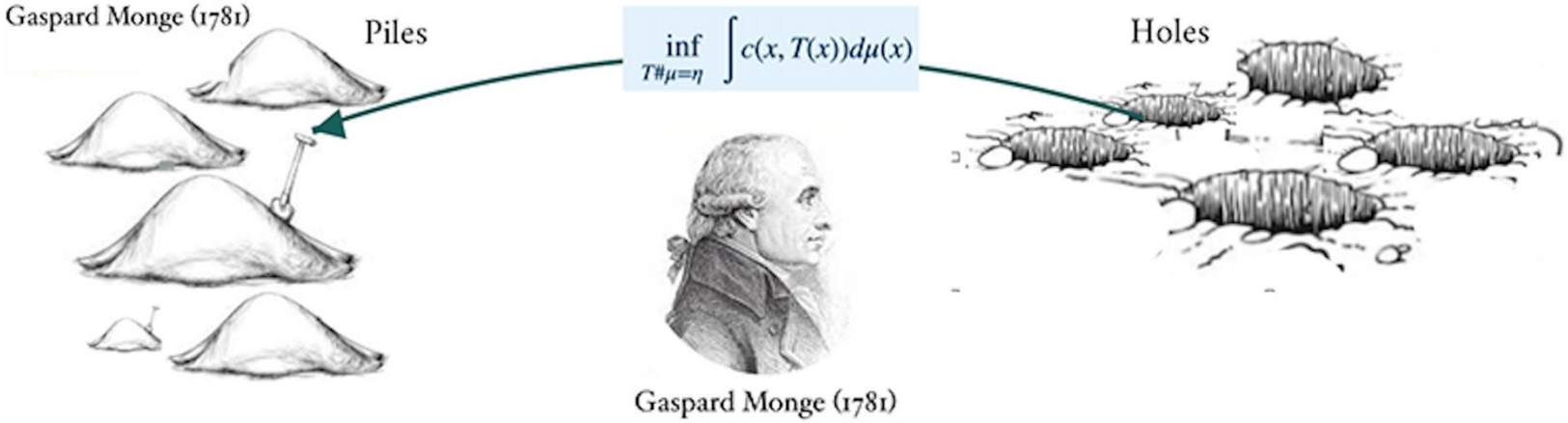
- Monge's Optimal Transport (1781)

  Given a margin (**r**, **c**) and cost matrix $\mathbf{C} = (c_{ij})$, c_{ij}=cost of moving a unit mass from location i to j, find the most efficient **map** $T: x \mapsto y$

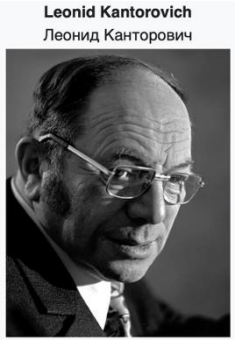


- Monge's Optimal Transport (1781)

  Given a margin (**r**, **c**) and cost matrix $\mathbf{C} = (c_{ij})$, c_{ij}=cost of moving a unit mass from location i to j, find the most efficient **map** $T: x \mapsto y$

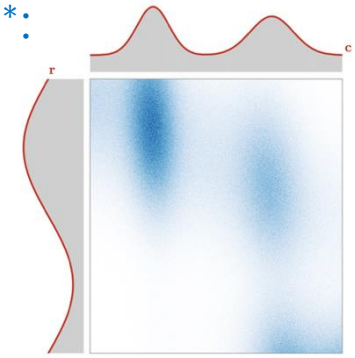


- Kantorovich's relaxation (1942)



In the same setting, find the most efficient **coupling** π^*:

$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle$$

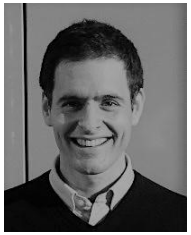


- Monge-Kantorovich Optimal Transport

$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle$$

- Monge-Kantorovich Optimal Transport

$$\pi^* = \operatorname*{arg\,min}_{\pi \in \Pi(\mathbf{r},\mathbf{c})} \langle \pi, \mathbf{C} \rangle$$



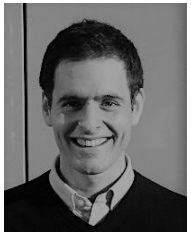
- Entropic Optimal Transport

$$\pi^* = \operatorname*{arg\,min}_{\pi \in \Pi(\mathbf{r},\mathbf{c})} \langle \pi, \mathbf{C} \rangle + \varepsilon \, \mathrm{KL}(\pi \| \mathbf{r} \otimes \mathbf{c})$$

- Entropic Regularization (Cuturi '13)

- Monge-Kantorovich Optimal Transport

$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle$$



- Entropic Optimal Transport

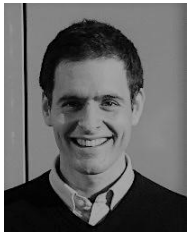
- Entropic Regularization (Cuturi '13)

$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle + \varepsilon \, \mathrm{KL}(\pi \| \mathbf{r} \otimes \mathbf{c})$$

$$= \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \left[D_{KL}(\pi \| \mathbf{A}) \text{ with } \mathbf{A} = \exp(-\mathbf{C}/\varepsilon) \odot (\mathbf{r} \otimes \mathbf{c}) \right]$$

(Matrix Scaling Problem, so solve it by Sinkhorn!)

- Monge-Kantorovich Optimal Transport



$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle$$

- Entropic Optimal Transport
  - Entropic Regularization (Cuturi '13)

$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle + \varepsilon \mathrm{KL}(\pi \| \mathbf{r} \otimes \mathbf{c})$$

$$= \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \left[D_{KL}(\pi \| \mathbf{A}) \text{ with } \mathbf{A} = \exp(-\mathbf{C}/\varepsilon) \odot (\mathbf{r} \otimes \mathbf{c}) \right]$$

(Matrix Scaling Problem, so solve it by Sinkhorn!)

- (Static) Schrodinger Bridge (1931-1932)

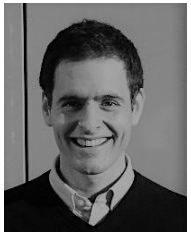


$$\min_{\mathscr{H} \in \Pi(\mu,\nu)} D_f(\mathscr{H} \| \mathscr{R}) \qquad \mathscr{R} = \text{Prior measure}$$

$$D_f(\mathscr{H} \| \mathscr{R}) := \int f\left(\frac{d\mathscr{H}}{d\mathscr{R}}\right) d\mathscr{R} \qquad \text{if } \mathscr{H} \ll \mathscr{R} \text{ and } \infty \text{ otherwise}$$

$$D_f = D_{KL} \text{ for } f(x) = x \log x$$

- Monge-Kantorovich Optimal Transport



$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle$$

- Entropic Optimal Transport
  - Entropic Regularization (Cuturi '13)

$$\pi^* = \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \langle \pi, \mathbf{C} \rangle + \varepsilon \mathrm{KL}(\pi \| \mathbf{r} \otimes \mathbf{c})$$

$$= \underset{\pi \in \Pi(\mathbf{r},\mathbf{c})}{\arg\min} \left[D_{KL}(\pi \| \mathbf{A}) \text{ with } \mathbf{A} = \exp(-\mathbf{C}/\varepsilon) \odot (\mathbf{r} \otimes \mathbf{c}) \right]$$

(Matrix Scaling Problem, so solve it by Sinkhorn!)

- (Static) Schrodinger Bridge (1931-1932)

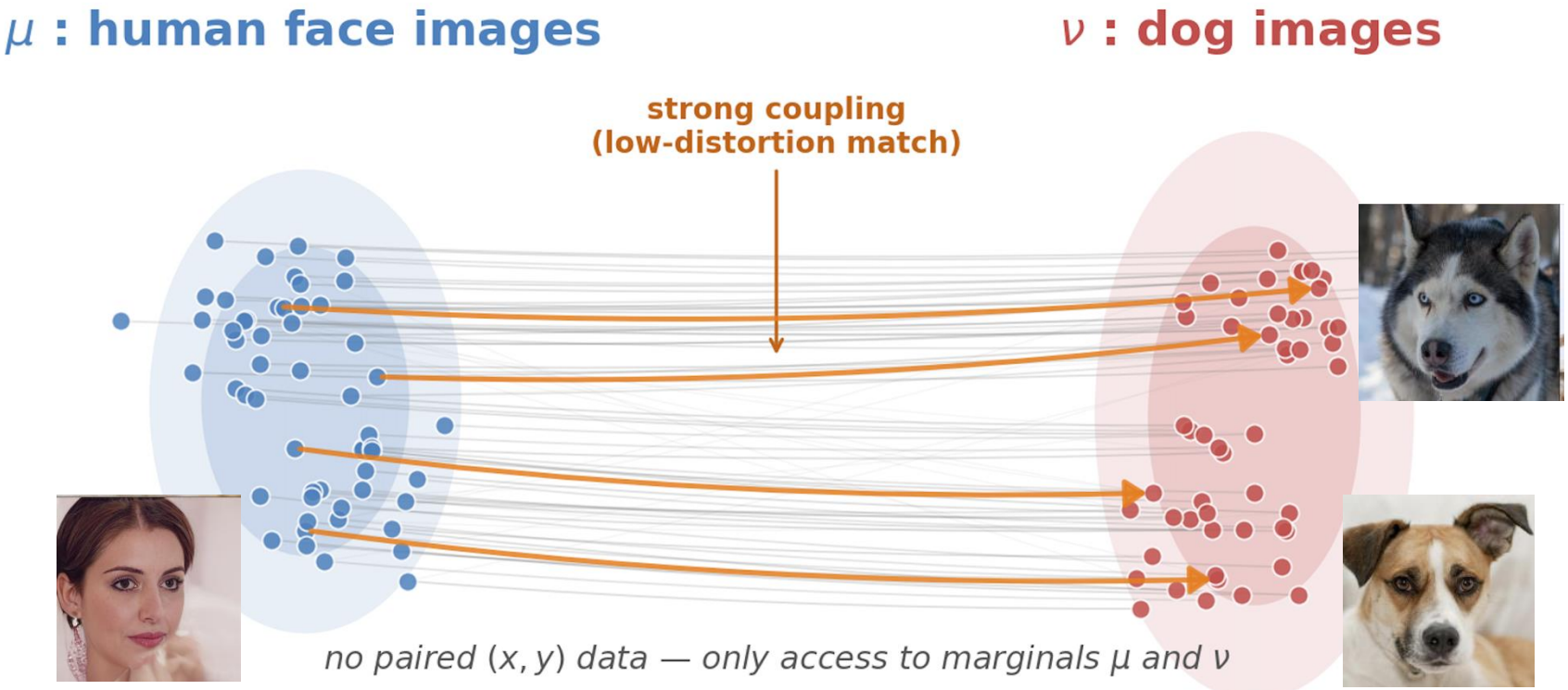


$$\min_{\mathscr{H} \in \Pi(\mu,\nu)} D_f(\mathscr{H} \| \mathscr{R}) \qquad \mathscr{R} = \text{Prior measure}$$

$$D_f(\mathscr{H} \| \mathscr{R}) := \int f\left(\frac{d\mathscr{H}}{d\mathscr{R}}\right) d\mathscr{R} \qquad \text{if } \mathscr{H} \ll \mathscr{R} \text{ and } \infty \text{ otherwise}$$

$$D_f = D_{KL} \text{ for } f(x) = x \log x$$

- Generative modeling: Unpaired Image Translation



- Generative modeling: Unpaired Image Translation



Results from REU 2025 @ UW-Madison on generative modeling

Dog ↔ Human image translation



Entropic transport plan π^*(dog, human) — concentrates on low-distortion matches

μ : dog image distribution

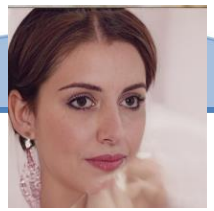
most likely dog image

$\pi^*(\cdot \mid y_{\text{human}})$ (conditional over dogs)



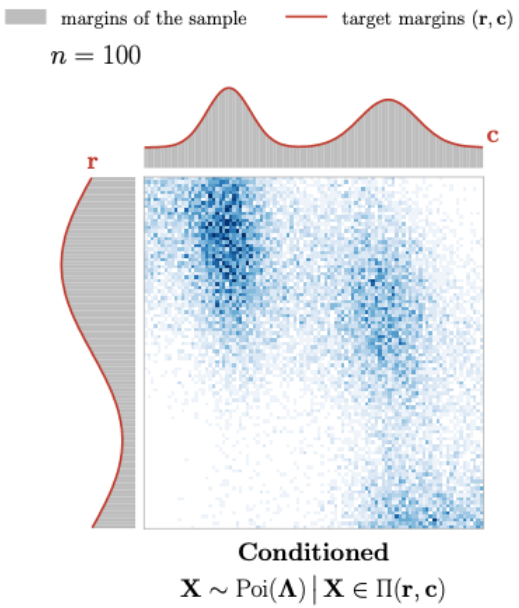
fixed human face y_{human}

ν : human face image distribution



Probabilistic Duality

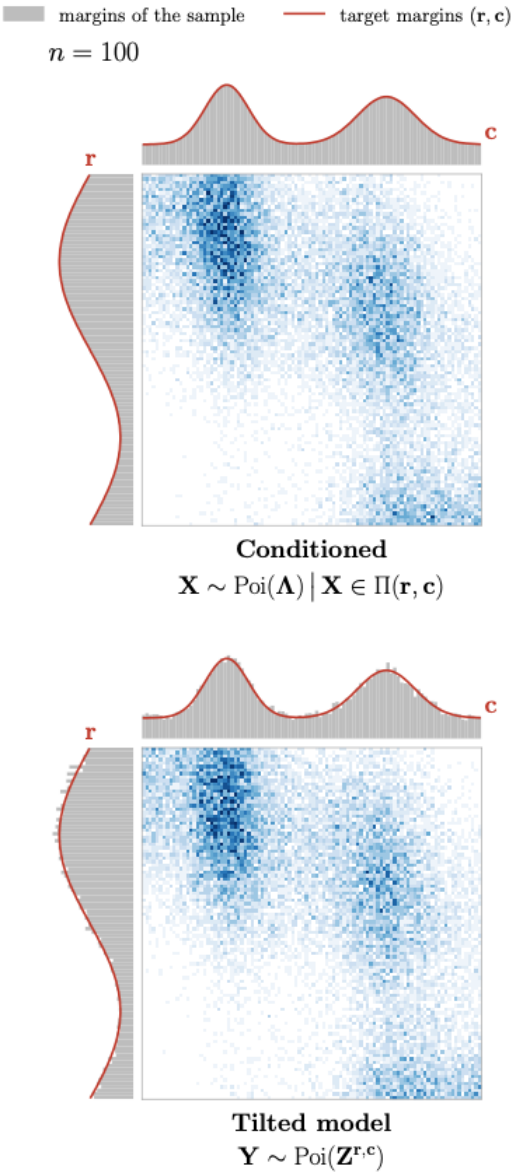
Poisson CT ≈ Tilted Poisson model



- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda_{ij}})$ entries) conditional on margin $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$

  Hard to analyze directly due to the induced dependence between the entries

Poisson CT ≈ Tilted Poisson model

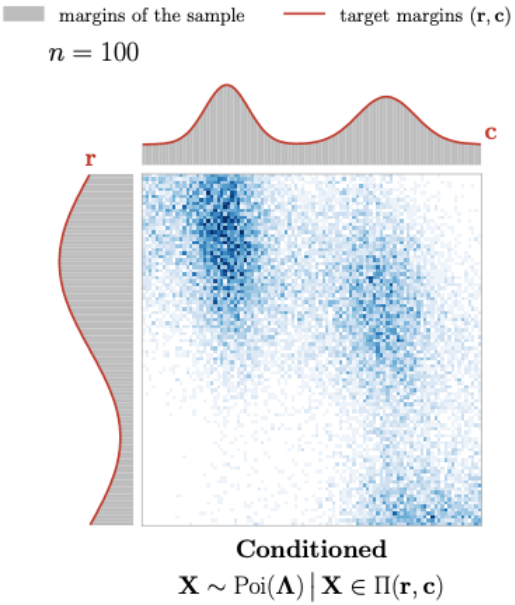


- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda_{ij}})$ entries) conditional on margin $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$

Hard to analyze directly due to the induced dependence between the entries

Q: Can we approximate it by $\boldsymbol{Y} \sim \text{Poisson}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some "optimal" mean $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$?

Poisson CT ≈ Tilted Poisson model



- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda_{ij}})$ entries) conditional on margin $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$

Hard to analyze directly due to the induced dependence between the entries

Q: Can we approximate it by $\boldsymbol{Y} \sim \text{Poisson}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some "optimal" mean $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$?

$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$ ← Relative Entropy Minimization

$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$ ← Maximum Likelihood Estimation

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{\boldsymbol{ij}})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method I) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some mean matrix $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{\boldsymbol{ij}})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method I) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some mean matrix $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$

Optimize $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$ via relative entropy minimization:

$$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}} = \arg\min_{\mathbf{Z} \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}(\text{Poisson}(\mathbf{Z}_{ij}) \,\|\, \text{Poisson}(\boldsymbol{\Lambda}_{ij}))$$

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{ij})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method I) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some mean matrix $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$

Optimize $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$ via relative entropy minimization:

$$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}} = \arg\min_{\mathbf{Z} \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}(\text{Poisson}(\mathbf{Z}_{ij}) \,\|\, \text{Poisson}(\boldsymbol{\Lambda}_{ij}))$$

$$\Updownarrow$$

$$\arg\min_{\mathbf{Z} \in \Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z} \| \mathbf{A}) = \sum_{ij} \mathbf{Z}_{ij} \log\left(\mathbf{Z}_{ij}/\mathbf{A}_{ij}\right) \right]$$

(Primal Problem: Matrix Scaling, SSB, EOT)

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{ij})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method I) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some mean matrix $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$

Optimize $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$ via relative entropy minimization:

$$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}} = \arg\min_{\mathbf{Z} \in \mathcal{T}(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}(\text{Poisson}(\mathbf{Z}_{ij}) \,\|\, \text{Poisson}(\boldsymbol{\Lambda}_{ij}))$$

$$\Updownarrow$$

$$\arg\min_{\mathbf{Z} \in \Pi(\mathbf{r},\mathbf{c})} \left[D_{KL}(\mathbf{Z} \| \mathbf{A}) = \sum_{ij} \mathbf{Z}_{ij} \log\left(\mathbf{Z}_{ij}/\mathbf{A}_{ij}\right) \right] \quad \textbf{(SSB)}$$

Thm. (Duan, L., Powell '26+) For bounded measurable functions h,

$$\mathbb{E}[h(\boldsymbol{X})] \leq (eN/n)^{2n}\, \mathbb{E}[h(\boldsymbol{Y})] \qquad N = total\ sum$$

Transference: "Events that are sufficiently rare under **Y** are also rare under **X**"

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{ij})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method II) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}))$ for unknown potential $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Optimize $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ via maximizing likelihood of observing margin $(\boldsymbol{r}, \boldsymbol{c})$

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{ij})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method II) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}))$ for unknown potential $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Optimize $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ via maximizing likelihood of observing margin $(\boldsymbol{r}, \boldsymbol{c})$

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \langle \boldsymbol{\Lambda}, \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \rangle \right] \quad \textbf{(Kantorovich Dual)}$$

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{ij})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method II) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}))$ for unknown potential $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Optimize $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ via maximizing likelihood of observing margin $(\boldsymbol{r}, \boldsymbol{c})$

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \langle \boldsymbol{\Lambda}, \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \rangle \right] \quad \textbf{(Kantorovich Dual)}$$

$$\Updownarrow$$

(MLE eq.) $\nabla g^{r,c} = 0 \Leftrightarrow \boldsymbol{Y} \sim \text{Poisson}(\boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}))$ satisfies $\mathbb{E}\boldsymbol{Y} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$

(Schrödinger System)

Poisson CT ≈ Tilted Poisson model

- $\boldsymbol{X} \sim \text{Poisson}(\boldsymbol{\Lambda})$ (independent $\text{Poisson}(\boldsymbol{\Lambda}_{ij})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- (Method II) Approximate $\boldsymbol{X}$ by $\boldsymbol{Y} \sim \text{Poisson}(\boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}))$ for unknown potential $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Optimize $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ via maximizing likelihood of observing margin $(\boldsymbol{r}, \boldsymbol{c})$

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[g^{\mathbf{r},\mathbf{c}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \mathbf{r}, \boldsymbol{\alpha} \rangle + \langle \mathbf{c}, \boldsymbol{\beta} \rangle - \langle \boldsymbol{\Lambda}, \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \rangle \right] \quad \textbf{(Kantorovich Dual)}$$

$$\Updownarrow$$

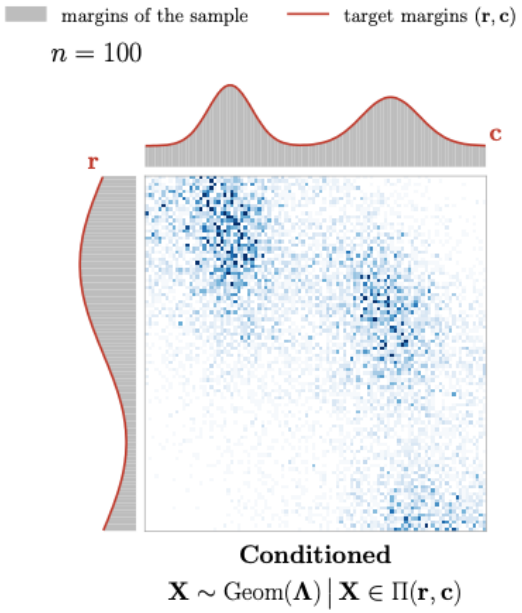
(MLE eq.) $\nabla g^{\boldsymbol{r},\boldsymbol{c}} = 0 \Leftrightarrow \boldsymbol{Y} \sim \text{Poisson}(\boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}))$ satisfies $\mathbb{E}\boldsymbol{Y} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$

(Schrödinger System)

- Compute MLE $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by standard Sinkhorn
- Methods I and II are in strong duality:

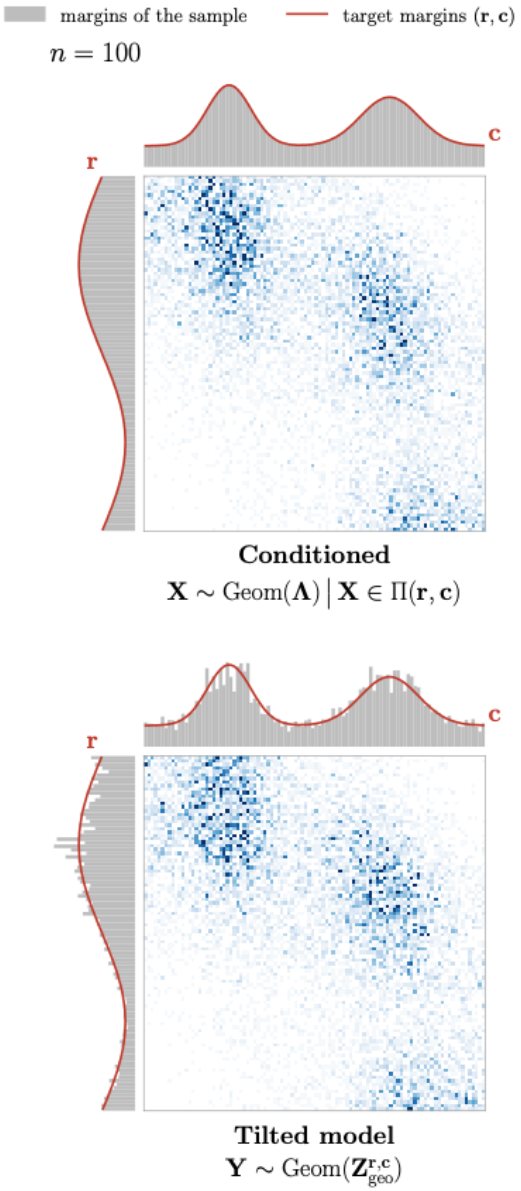
$$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}} = \boldsymbol{\Lambda} \odot \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta})$$

Geometric CT ≈ Tilted Geometric model



- $\boldsymbol{X} \sim \mathrm{Geom}(\boldsymbol{\Lambda})$ (independent $\mathrm{Geom}\left(\mathrm{mean} = \boldsymbol{\Lambda}_{\boldsymbol{ij}}\right)$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- $\boldsymbol{Y} \sim \mathrm{Geom}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some "optimal" mean $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$

Geometric CT ≈ Tilted Geometric model



- $\boldsymbol{X} \sim \mathrm{Geom}(\boldsymbol{\Lambda})$ (independent $\mathrm{Geom}(\mathrm{mean} = \boldsymbol{\Lambda_{ij}})$ entries) conditional on $\boldsymbol{X} \in \Pi(\boldsymbol{r}, \boldsymbol{c})$
- $\boldsymbol{Y} \sim \mathrm{Geom}(\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}})$ for some "optimal" mean $\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$

$$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}} = \operatorname*{argmax}_{\mathbf{Z} \in \Pi(\boldsymbol{r},\boldsymbol{c})} \sum_{i,j} \underbrace{D_{KL}(\mathrm{Geom}(\mathbf{Z}_{ij}) \,\|\, \mathrm{Geom}(\mathbf{\Lambda}_{ij}))}_{=(1+\mathbf{Z}_{ij})\log(1+\mathbf{Z}_{ij}) - \mathbf{Z}_{ij}\log\mathbf{Z}_{ij} + \mathbf{Z}_{ij}\log\frac{\mathbf{\Lambda}_{ij}}{1+\mathbf{\Lambda}_{ij}}}$$



Alexander Barvinok

$\mathbf{Z}^{\boldsymbol{r},\boldsymbol{c}}$ (for $\boldsymbol{\Lambda_{ij}} \equiv 1$) known as the "typical table" due to Barvinok (2008)

General Probabilistic Kantorovich Duality

Thm. (L., Mukherjee '25+)

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

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}\left(\mu_{\phi(\mathbf{Z}_{ij})} \,\|\, \mu_{\phi(\boldsymbol{\Lambda}_{ij})}\right) = \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{ij}\left[\psi\left(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\phi(\boldsymbol{\Lambda}_{ij})\right) - \psi\left(\phi(\boldsymbol{\Lambda}_{ij})\right)\right]$$

$$\mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \psi'(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\phi(\boldsymbol{\Lambda}_{ij})) \qquad \text{(Optimal Solution)}$$

General Probabilistic Kantorovich Duality

Thm. (L., Mukherjee '25+)

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

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}\left(\mu_{\phi(\mathbf{Z}_{ij})} \,\|\, \mu_{\phi(\mathbf{\Lambda}_{ij})}\right) = \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ \langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{ij}\left[\psi\left(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\phi(\mathbf{\Lambda}_{ij})\right) - \psi\left(\phi(\mathbf{\Lambda}_{ij})\right)\right]$$

$$\mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \psi'(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\phi(\mathbf{\Lambda}_{ij})) \qquad \text{(Optimal Solution)}$$

$(\mathbf{r},\mathbf{c}) = m \times n$ margin

$\mathbf{\Lambda} = (\mathbf{\Lambda}_{ij}) = m \times n$ prior mean matrix

$\mu =$ base probability distribution

$$\mu = \text{Poisson}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \exp(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j)\mathbf{\Lambda}_{ij}$$

$$\mu = \text{Geometric}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \frac{1}{\exp(-\boldsymbol{\alpha}(i)-\boldsymbol{\beta}(j)+\log(1+\mathbf{\Lambda}^{-1}_{ij}))-1}$$

General Probabilistic Kantorovich Duality

Thm. (L., Mukherjee '25+)

$$\arg\min_{Z} \ \underline{\textit{Relative Entropy}}(Z) \Leftrightarrow \arg\max_{\alpha,\beta} \ \underline{\textit{Likelihood Of Margin}}(\alpha,\beta)$$

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}\left(\mu_{\phi(\mathbf{Z}_{ij})} \,\|\, \mu_{\phi(\mathbf{\Lambda}_{ij})}\right) = \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{ij}\left[\psi\left(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\phi(\mathbf{\Lambda}_{ij})\right) - \psi\left(\phi(\mathbf{\Lambda}_{ij})\right)\right]$$

$$\mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \psi'(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\phi(\mathbf{\Lambda}_{ij})) \qquad \text{(Optimal Solution)}$$

$$\mu = \text{Poisson}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \exp(\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j)\mathbf{\Lambda}_{ij}$$

$$\mu = \text{Geometric}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \frac{1}{\exp(-\boldsymbol{\alpha}(i)-\boldsymbol{\beta}(j)+\log(1+\mathbf{\Lambda}_{ij}^{-1}))-1}$$

$(\mathbf{r},\mathbf{c}) = m\times n$ margin

$\mathbf{\Lambda} = (\mathbf{\Lambda}_{ij}) = m\times n$ prior mean matrix

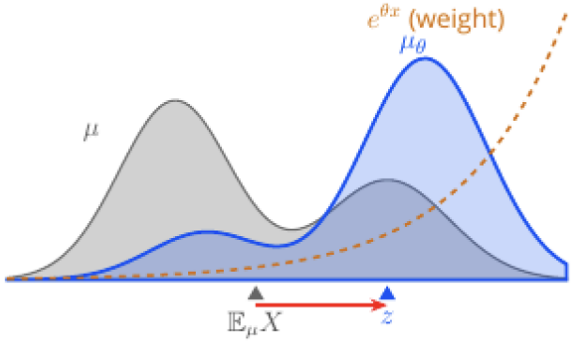
$\mu =$ base probability distribution

$\psi(\theta) = \log\int e^{\theta x}\, d\mu(x) =$ log MGF of μ

$\mu_\theta =$ expoential tilt of μ: $\dfrac{d\mu_\theta}{d\mu}(x) = e^{\theta x-\psi(\theta)}$

$\psi'(\theta) =$ mean of the tilted distribution μ_θ

$\phi = (\psi')^{-1} =$ mean-to-tild conversion



Exponential tilting

General Probabilistic Kantorovich Duality

Thm. (L., Mukherjee '25+)

$$\arg\min_{Z} \ \underline{\textit{Relative Entropy}}(Z) \Leftrightarrow \arg\max_{\alpha,\beta} \ \underline{\textit{Likelihood Of Margin}}(\alpha,\beta)$$

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}\left(\mu_{\phi(\mathbf{Z}_{ij})} \,\|\, \mu_{\phi(\mathbf{\Lambda}_{ij})}\right) = \sup_{\alpha,\beta} \langle \mathbf{r},\alpha\rangle + \langle \mathbf{c},\beta\rangle - \sum_{ij}\left[\psi\left(\alpha_i+\beta_j+\phi(\mathbf{\Lambda}_{ij})\right) - \psi\left(\phi(\mathbf{\Lambda}_{ij})\right)\right]$$

$$\mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \psi'(\alpha_i+\beta_j+\phi(\mathbf{\Lambda}_{ij})) \qquad \text{(Optimal Solution)}$$

$$\mu = \text{Poisson}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \exp(\alpha_i+\beta_j)\mathbf{\Lambda}_{ij}$$

$$\mu = \text{Geometric}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \frac{1}{\exp(-\alpha(i)-\beta(j)+\log(1+\mathbf{\Lambda}_{ij}^{-1}))-1}$$

$(\mathbf{r},\mathbf{c}) = m\times n$ margin

$\mathbf{\Lambda} = (\mathbf{\Lambda}_{ij}) = m\times n$ prior mean matrix

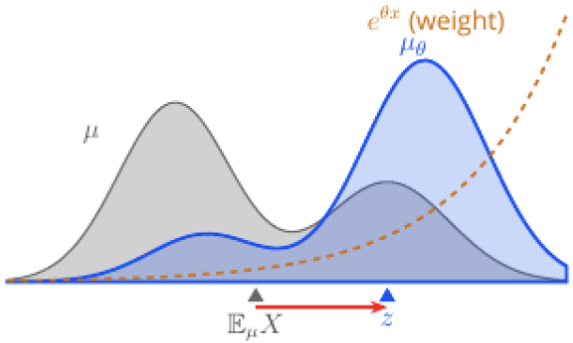
$\mu =$ base probability distribution

$\psi(\theta) = \log\int e^{\theta x}\, d\mu(x) =$ log MGF of μ

$\mu_\theta =$ expoential tilt of μ: $\dfrac{d\mu_\theta}{d\mu}(x) = e^{\theta x-\psi(\theta)}$

$\psi'(\theta) =$ mean of the tilted distribution μ_θ

$\phi = (\psi')^{-1} =$ mean-to-tild conversion

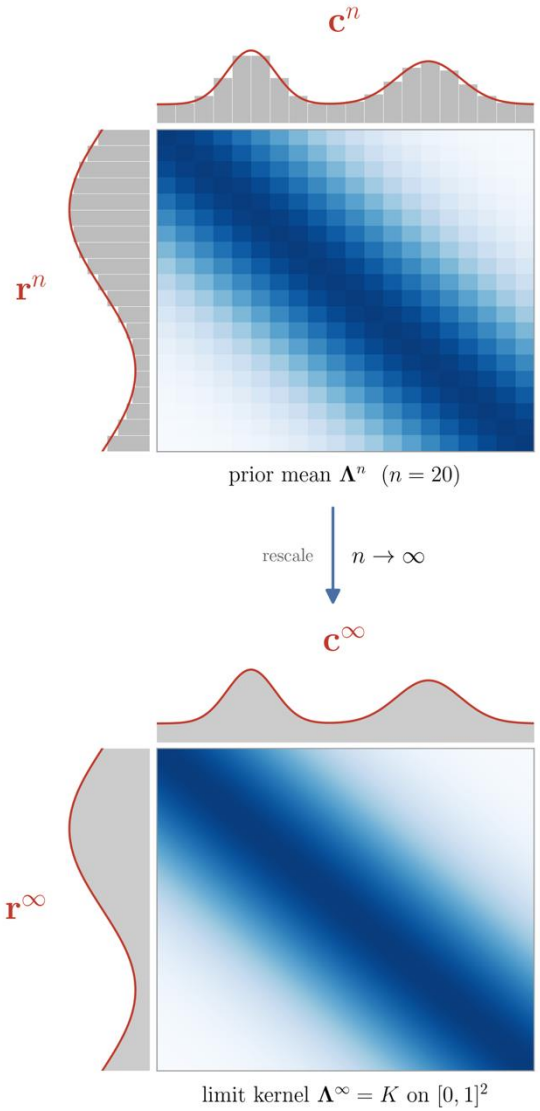


Exponential tilting

Our next goal is "scaling limit": $m,n\to\infty$

Scaling limit

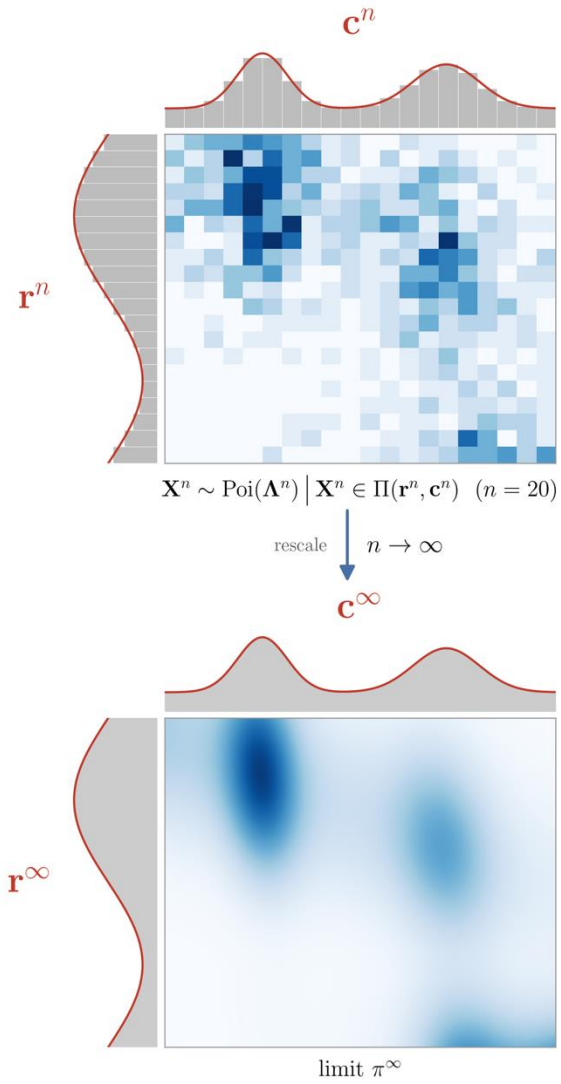
Our next goal is "scaling limit": matrix dimension $n \to \infty$



Assume margin vectors & Prior mean matrices converge in scaling limit

$$(\boldsymbol{r^n}, \boldsymbol{c^n}, \boldsymbol{\Lambda^n}) \to (\boldsymbol{r^\infty}, \boldsymbol{c^\infty}, \boldsymbol{\Lambda^\infty})$$

Our next goal is "scaling limit": matrix dimension $n \to \infty$

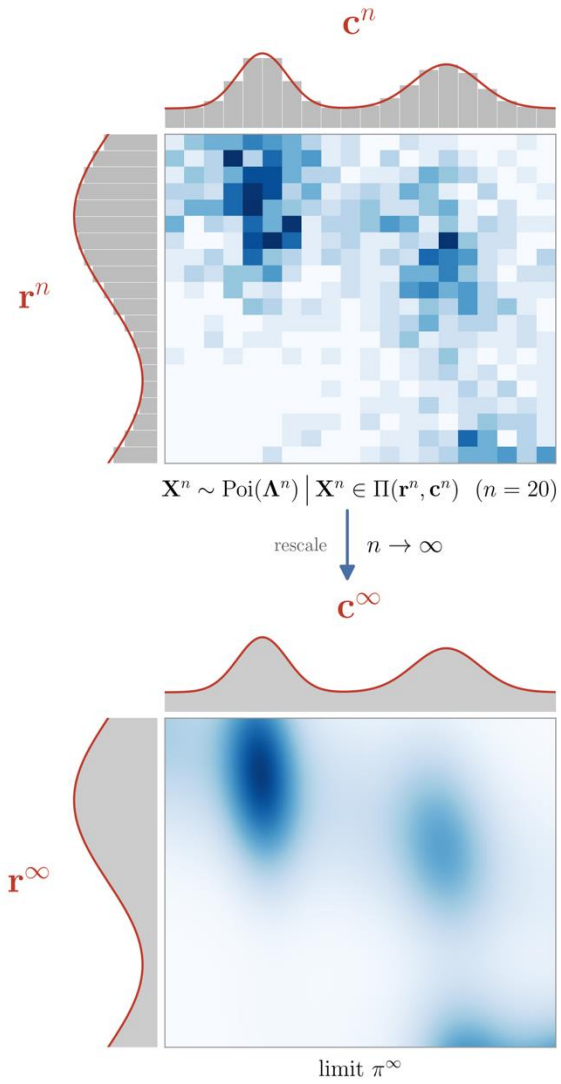


$$\approx \quad \mathbf{Z}^{\mathbf{r}^n,\mathbf{c}^n} = \operatorname*{arg\,min}_{\mathbf{Z}\in\Pi(\mathbf{r}^n,\mathbf{c}^n)} \sum_{ij} D_{KL}\left(\mu_{\phi(\mathbf{Z}_{ij})} \,\|\, \mu_{\phi(\mathbf{\Lambda}^n_{ij})}\right)$$

?

$$= \quad \pi^\infty = \operatorname*{arg\,min}_{\pi\in\Pi(\mathbf{r}^\infty,\mathbf{c}^\infty)} \left[\cdots\right]$$

Our next goal is "scaling limit": matrix dimension $n \to \infty$



$$\approx \quad \mu = \text{Poisson}: \quad \mathbf{Z}_{ij}^{\mathbf{r},\mathbf{c}} = \exp(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_j)\mathbf{\Lambda}_{ij}$$

$$\mu = \text{Geometric}: \quad \mathbf{Z}_{ij}^{\mathbf{r},\mathbf{c}} = \frac{1}{\exp(-\boldsymbol{\alpha}(i) - \boldsymbol{\beta}(j) + \log(1 + \mathbf{\Lambda}_{ij}^{-1})) - 1}$$

?

$$= \quad \pi^\infty = \underset{\pi \in \Pi(\mathbf{r}^\infty, \mathbf{c}^\infty)}{\arg\min} \; [\cdots]$$

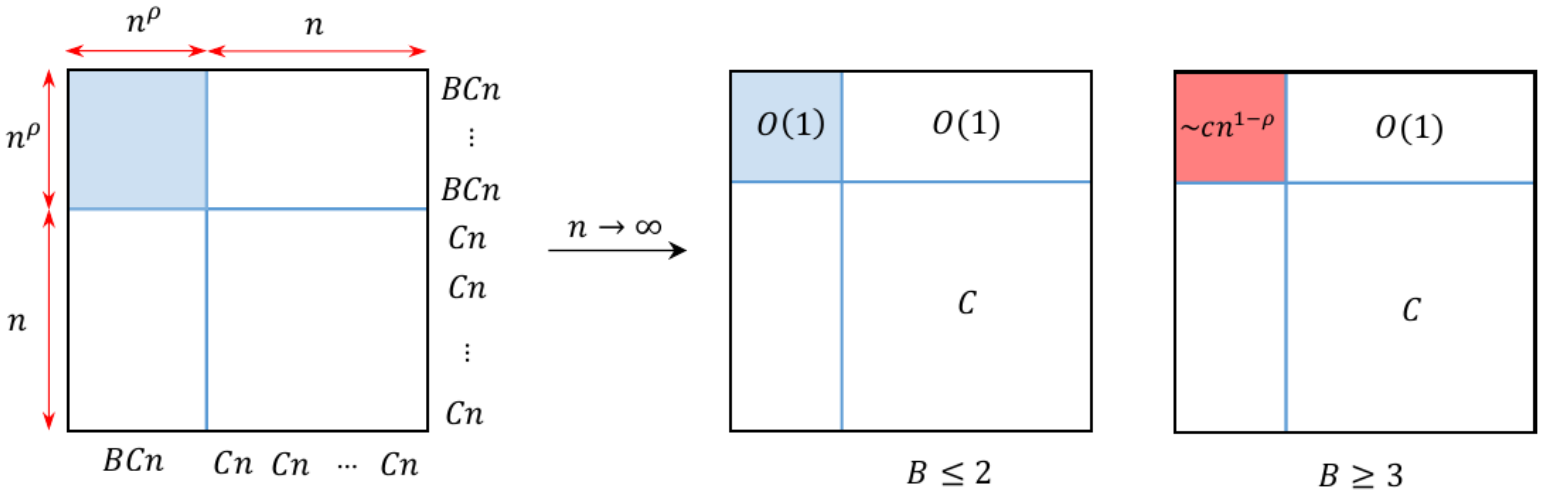
Barvinok's typical table ($\mu = Geometric$)

- For a $m \times n$ margin $(\mathbf{r}, \mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \underset{Z \in \mathcal{T}(\mathbf{r},\mathbf{c})}{\arg\max} \left[g(Z) = \sum_{i,j} (z_{ij}+1)\log(z_{ij}+1) - z_{ij}\log(z_{ij}) \right]$$

$$\mu = \text{Geometric}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \frac{1}{\exp(-\boldsymbol{\alpha}(i) - \boldsymbol{\beta}(j)) - 1}$$

- How does $Z^{r,c}$ depend on the margin $(\mathbf{r},\mathbf{c})$? Does it always stay bounded?
- In 2010, Barbinok conjectured that there is a phase transition in Uniform($\mathcal{T}$(Barv. margin)) as B increases using typical tables



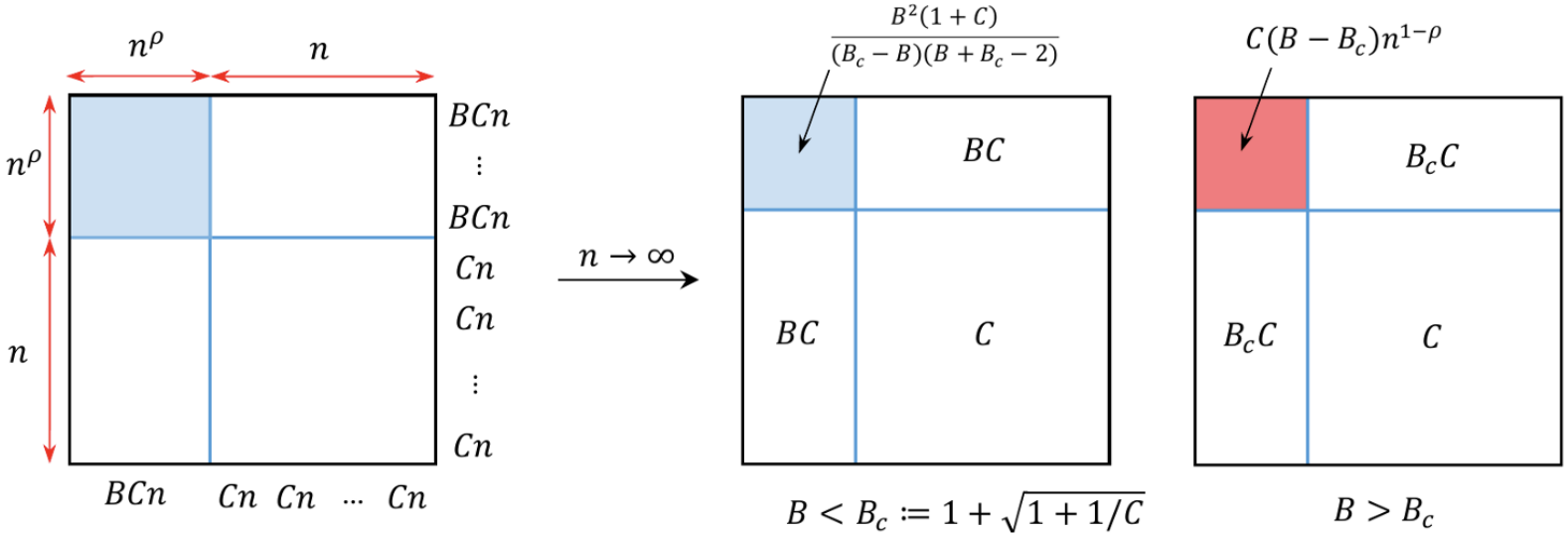
Barvinok's typical table ($\mu = Geometric$)

- For a $m \times n$ margin $(\mathbf{r}, \mathbf{c})$, the corresponding typical table is

$$Z^{\mathbf{r},\mathbf{c}} := \underset{Z \in \mathcal{T}(\mathbf{r},\mathbf{c})}{\arg\max} \left[g(Z) = \sum_{i,j} (z_{ij}+1)\log(z_{ij}+1) - z_{ij}\log(z_{ij}) \right]$$

$$\mu = \text{Geometric}: \quad \mathbf{Z}^{\mathbf{r},\mathbf{c}}_{ij} = \frac{1}{\exp(-\boldsymbol{\alpha}(i) - \boldsymbol{\beta}(j)) - 1}$$

- How does $Z^{r,c}$ depend on the margin $(\mathbf{r},\mathbf{c})$? Does it always stay bounded?
- [Dittmer, L., Pak '20]: Sharp phase transition in typical tables for Barvinok margin



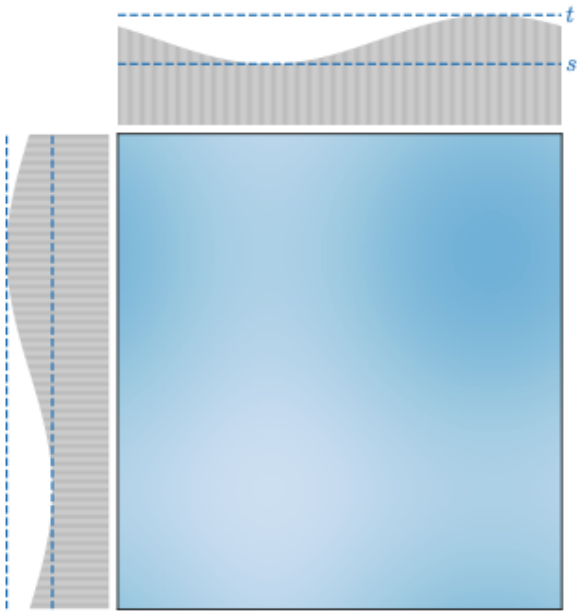
Barvinok's typical table ($\mu = Geometric$)

Thm. (L., Mukherjee '25+) If $s \le \frac{r}{n} \le t$ and $s \le \frac{c}{m} \le t$ entrywise, then $\boldsymbol{Z^{r,c}}$ is uniformly bounded by M=M(s,t) provided

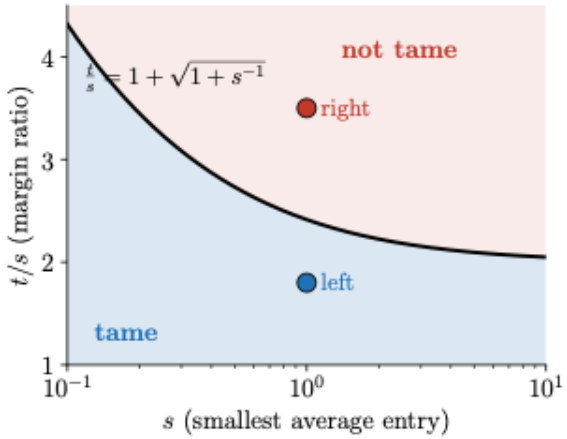
$$t/s < 1 + \sqrt{1 + s^{-1}}$$

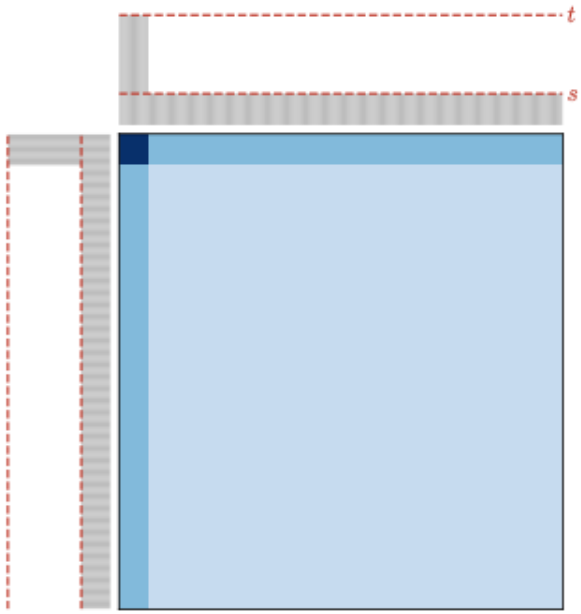
This result is sharp due the phase transition by [Dittmer, L., Pak. '20]

Typical tables (geometric entries, $n \approx 200$): margins within a factor $\lambda_c(s)$ of each other $\Rightarrow$ tame



similar-looking margins: $t/s = 1.8 < \lambda_c$
tame: $\max_{ij} z_{ij} = 2.4$





different-looking margins: $t/s = 3.5 > \lambda_c$
not tame: corner $z_{11} = 22$ (grows like $\sqrt{n}$)

Stability of the primal and dual solutions (Λ = homogeneous)

General μ, constant prior mean $\mathbf{\Lambda}_{ij}$:

$$\min_{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})} \sum_{ij} D_{KL}\left(\mu_{\phi(\mathbf{Z}_{ij})} \,\|\, \mu\right) = \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \mathbf{r}, \boldsymbol{\alpha}\rangle + \langle \mathbf{c}, \boldsymbol{\beta}\rangle - \sum_{ij} \psi\left(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_j\right)$$

How much does the rescaled matrix / potentials change when we perturb the margin $(\boldsymbol{r}, \boldsymbol{c})$?

Thm. (L., Mukherjee 25+)

Two "$\delta - tame$" continuum margins $(\mathbf{r}, \mathbf{c}), (\mathbf{r}', \mathbf{c}')$:

$$\|W^{\mathbf{r},\mathbf{c}} - W^{\mathbf{r}',\mathbf{c}'}\|_2^2 \le 2C_\delta \left(\sup_{\phi(A_\delta)\le w\le \phi(B_\delta)} \psi''(w) \right) \|(\mathbf{r},\mathbf{c}) - (\mathbf{r}',\mathbf{c}')\|_1.$$

Cor. Convergence of margins within tame regime → Scaling limit

Stability of the primal and dual solutions (μ = Poisson, Λ = general)

Poisson μ, general prior mean $\mathbf{\Lambda}_{ij}$:

$$\mathbf{Z}^{\mathbf{r},\mathbf{c}} = \underset{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})}{\arg\min}\, D_{KL}(\mathbf{Z} \,\|\, \mathbf{\Lambda}) = \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) \odot \mathbf{\Lambda}$$

Thm. (Duan, Powell, L. 26a+)

$$d_{\text{Hellinger}}(\Lambda^{r,c}, \Sigma^{r,c}) \le C\, d_{\text{Hellinger}}(\Lambda, \Sigma)$$

$$d_{\text{Hellinger}}\left(\Lambda^{\boldsymbol{r},\boldsymbol{c}}, \Lambda^{\boldsymbol{r}',\boldsymbol{c}'}\right) \le C' \|(\boldsymbol{r},\boldsymbol{c}) - (\boldsymbol{r}',\boldsymbol{c}')\|_1$$

$$\|\left(\boldsymbol{\alpha}_{\Lambda}^{\boldsymbol{r},\boldsymbol{c}}, \boldsymbol{\beta}_{\Lambda}^{\boldsymbol{r},\boldsymbol{c}}\right) - \left(\boldsymbol{\alpha}_{\Lambda}^{\boldsymbol{r}',\boldsymbol{c}'}, \boldsymbol{\beta}_{\Lambda}^{\boldsymbol{r}',\boldsymbol{c}'}\right)\|_\infty \le C'' \text{(relative margin error)}$$

Cor. Convergence of margins & Prior means → Scaling limit

Stability of the primal and dual solutions (μ = Poisson, Λ = general)

Poisson μ, general prior mean $\boldsymbol{\Lambda}_{ij}$:

$$\mathbf{Z}^{\mathbf{r},\mathbf{c}} = \underset{\mathbf{Z}\in\Pi(\mathbf{r},\mathbf{c})}{\arg\min}\, D_{KL}(\mathbf{Z}\,\|\,\boldsymbol{\Lambda}) = \exp(\boldsymbol{\alpha}\oplus\boldsymbol{\beta})\odot\boldsymbol{\Lambda}$$

Thm. (Duan, Powell, L. 26a+)

$$d_{\text{Hellinger}}(\Lambda^{r,c},\Sigma^{r,c}) \le C\, d_{\text{Hellinger}}(\Lambda,\Sigma)$$

$$d_{\text{Hellinger}}\left(\Lambda^{\boldsymbol{r},\boldsymbol{c}},\Lambda^{\boldsymbol{r}',\boldsymbol{c}'}\right) \le C'\|(\boldsymbol{r},\boldsymbol{c})-(\boldsymbol{r}',\boldsymbol{c}')\|_1$$

$$\|\left(\boldsymbol{\alpha}_{\Lambda}^{\boldsymbol{r},\boldsymbol{c}},\boldsymbol{\beta}_{\Lambda}^{\boldsymbol{r},\boldsymbol{c}}\right)-\left(\boldsymbol{\alpha}_{\Lambda}^{\boldsymbol{r}',\boldsymbol{c}'},\boldsymbol{\beta}_{\Lambda}^{\boldsymbol{r}',\boldsymbol{c}'}\right)\|_\infty \le C''\,(\text{relative margin error})$$

Cor. Convergence of margins & Prior means → Scaling limit

c.f. No issue of phase transition in the Poisson case:

$$\|(\alpha^*,\beta^*)\|_\infty \le \frac{3}{2}\|\kappa\|_{max} \qquad \kappa_{ij} = -\log\frac{\boldsymbol{\Lambda}_{ij}\boldsymbol{N}^2}{\|\boldsymbol{\Lambda}\|_1\boldsymbol{r}_i\,\boldsymbol{c}_j}$$

Optimization

Dual Problem

- General formulation

$$\sup_{\alpha,\beta}\left[f(\alpha,\beta)=\langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\left[\psi\left(\alpha_i+\beta_j+\phi(\Lambda_{ij})\right)-\psi\left(\phi(\Lambda_{ij})\right)\right]\right]$$

$$\sup_{\alpha,\beta}\ \langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\exp(\alpha_i+\beta_j)\Lambda_{ij} \qquad (\mu=\text{Poisson; EOT dual})$$

$$\sup_{\alpha,\beta}\ \langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\Lambda_{ij}\left(e^{\alpha_i+\beta_j}-1\right) \qquad (\mu=\text{Geometric; Typical table dual})$$

Sinkhorn Algorithm

$$\begin{cases}\forall 1\le i\le n,\ \beta_k(j)\leftarrow\log\frac{\mathbf{c}(j)}{\sum_{i=1}^m \mathbf{A}_{ij}\exp(\alpha_{k-1}(i))},\\ \forall 1\le i\le m,\ \alpha_k(i)\leftarrow\log\frac{\mathbf{r}(i)}{\sum_{j=1}^n \mathbf{A}_{ij}\exp(\beta_k(j))}.\end{cases}$$

$=$

Alternating maximization

$$\begin{cases}\beta_k\leftarrow\arg\max_\beta f(\alpha_{k-1},\beta)\\ \alpha_k\leftarrow\arg\max_\alpha f(\alpha,\beta_k)\end{cases}$$

- Sticky (interesting) points for optimization:
  Lipschitz constant L, strong convexity μ of the dual objective are unbounded

→ Need a priori control on the range of the iterates (α_k,β_k)

Dual Problem

- General formulation

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[f(\boldsymbol{\alpha},\boldsymbol{\beta}) = \langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{ij} \left[\psi\left(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \phi(\boldsymbol{\Lambda}_{ij})\right) - \psi\left(\phi(\boldsymbol{\Lambda}_{ij})\right) \right] \right]$$

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{ij} \exp(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_j)\boldsymbol{\Lambda}_{ij} \qquad (\mu = \text{Poisson; EOT dual})$$

$$\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \mathbf{r},\boldsymbol{\alpha}\rangle + \langle \mathbf{c},\boldsymbol{\beta}\rangle - \sum_{ij} \boldsymbol{\Lambda}_{ij}\left(e^{\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j} - 1\right) \qquad (\mu = \text{Geometric; Typical table dual})$$

Dual Problem

- General formulation

$$\sup_{\alpha,\beta}\left[f(\alpha,\beta)=\langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\left[\psi\left(\alpha_i+\beta_j+\phi(\Lambda_{ij})\right)-\psi\left(\phi(\Lambda_{ij})\right)\right]\right]$$

$$\sup_{\alpha,\beta}\ \langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\exp(\alpha_i+\beta_j)\Lambda_{ij} \qquad (\mu=\text{Poisson; EOT dual})$$

$$\sup_{\alpha,\beta}\ \langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\Lambda_{ij}\left(e^{\alpha_i+\beta_j}-1\right) \qquad (\mu=\text{Geometric; Typical table dual})$$

Sinkhorn Algorithm

$$\begin{cases}\forall 1\le i\le n,\ \beta_k(j)\leftarrow\log\frac{\mathbf{c}(j)}{\sum_{i=1}^m \mathbf{A}_{ij}\exp(\alpha_{k-1}(i))},\\ \forall 1\le i\le m,\ \alpha_k(i)\leftarrow\log\frac{\mathbf{r}(i)}{\sum_{j=1}^n \mathbf{A}_{ij}\exp(\beta_k(j))}.\end{cases}$$

$=$

Alternating maximization

$$\begin{cases}\beta_k\leftarrow\arg\max_\beta f(\alpha_{k-1},\beta)\\ \alpha_k\leftarrow\arg\max_\alpha f(\alpha,\beta_k)\end{cases}$$

Dual Problem

- General formulation

$$\sup_{\alpha,\beta}\left[f(\alpha,\beta)=\langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\left[\psi\left(\alpha_i+\beta_j+\phi(\Lambda_{ij})\right)-\psi\left(\phi(\Lambda_{ij})\right)\right]\right]$$

$$\sup_{\alpha,\beta}\ \langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\exp(\alpha_i+\beta_j)\Lambda_{ij} \qquad (\mu=\text{Poisson; EOT dual})$$

$$\sup_{\alpha,\beta}\ \langle \mathbf{r},\alpha\rangle+\langle \mathbf{c},\beta\rangle-\sum_{ij}\Lambda_{ij}\left(e^{\alpha_i+\beta_j}-1\right) \qquad (\mu=\text{Geometric; Typical table dual})$$

Sinkhorn Algorithm

$$\begin{cases}\forall 1\le i\le n,\ \beta_k(j)\leftarrow\log\frac{\mathbf{c}(j)}{\sum_{i=1}^m \mathbf{A}_{ij}\exp(\alpha_{k-1}(i))},\\ \forall 1\le i\le m,\ \alpha_k(i)\leftarrow\log\frac{\mathbf{r}(i)}{\sum_{j=1}^n \mathbf{A}_{ij}\exp(\beta_k(j))}.\end{cases}$$

$=$

Alternating maximization

$$\begin{cases}\beta_k\leftarrow\arg\max_\beta f(\alpha_{k-1},\beta)\\ \alpha_k\leftarrow\arg\max_\alpha f(\alpha,\beta_k)\end{cases}$$

- Sticky (interesting) points for optimization:
  Lipschitz constant L, strong convexity μ of the dual objective are unbounded

→ Need a priori control on the range of the iterates (α_k,β_k)

- Quick math for the convergence of Sinkhorn's Algorithm
  (Dual interpretation of Franklin and Lorenz '89)

Step 1: ℓ_∞- non-expansion of the potentials

$$\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|_\infty = \|\xi(\boldsymbol{\beta}_{k-1}) - \xi(\boldsymbol{\beta}^*)\|_\infty \qquad (\ \xi : \beta \mapsto \alpha \text{ update map}\ )$$

$$= \left\| \int_0^1 J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})\, ds (\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*) \right\|_\infty$$

- Quick math for the convergence of Sinkhorn's Algorithm
  (Dual interpretation of Franklin and Lorenz '89)

Step 1: ℓ_∞- non-expansion of the potentials

$$\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|_\infty = \|\xi(\boldsymbol{\beta}_{k-1}) - \xi(\boldsymbol{\beta}^*)\|_\infty \qquad (\xi: \beta \mapsto \alpha \text{ update map})$$

$$= \left\| \int_0^1 J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})\, ds (\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*) \right\|_\infty$$

$$= \left(\int_0^1 \| \underbrace{J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})}_{=\text{row-stochastic}} \|_{\infty\to\infty} ds \right) \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_\infty$$

$$- \begin{bmatrix} 0 & [E_{ij}/E_{i\bullet}]_{m\times n} \\ [E_{ij}/E_{\bullet j}]^\top_{n\times m} & 0 \end{bmatrix}$$

- Quick math for the convergence of Sinkhorn's Algorithm
  (Dual interpretation of Franklin and Lorenz '89)

Step 1: ℓ_∞- non-expansion of the potentials

$$\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|_\infty = \|\xi(\boldsymbol{\beta}_{k-1}) - \xi(\boldsymbol{\beta}^*)\|_\infty \qquad (\xi: \beta \mapsto \alpha \text{ update map})$$

$$= \left\| \int_0^1 J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})\, ds (\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*) \right\|_\infty$$

$$= \left(\int_0^1 \| \underbrace{J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})}_{=\text{row-stochastic}} \|_{\infty\to\infty} ds \right) \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_\infty$$

$$\le \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_\infty$$

$$\le \cdots \le \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^*\|_\infty$$

$$- \begin{bmatrix} 0 & [E_{ij}/E_{i\bullet}]_{m\times n} \\ [E_{ij}/E_{\bullet j}]^\top_{n\times m} & 0 \end{bmatrix}$$

- Quick math for the convergence of Sinkhorn's Algorithm
  (Dual interpretation of Franklin and Lorenz '89)

Step 2: span-contraction of the potentials

$$\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|_{sp} = \|\xi(\boldsymbol{\beta}_{k-1}) - \xi(\boldsymbol{\beta}^*)\|_{sp} \qquad (\, sp(x) := \max x \, - \min x \,)$$

$$= \left\| \int_0^1 J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})\, ds (\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*) \right\|_{sp}$$

$$= \Big(\int_0^1 \Big\| \underbrace{J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})}_{\text{=row-stochastic, nonzero}} \Big\|_{sp\to sp} ds \Big) \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_{sp}$$

$\leq \tanh(\Delta(J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s}))/2) < 1$ (Birkoff's contraction Thm.)

$$\Delta(P) := \max_{i,i',j,j'} \log \frac{P_{ij}\, P_{i'j'}}{P_{ij'}\, P_{i'j}}$$

- Quick math for the convergence of Sinkhorn's Algorithm
  (Dual interpretation of Franklin and Lorenz '89)

Step 2: span-contraction of the potentials

$$\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|_{sp} = \|\xi(\boldsymbol{\beta}_{k-1}) - \xi(\boldsymbol{\beta}^*)\|_{sp} \qquad (\,sp(x) := \max x - \min x\,)$$

$$= \left\| \int_0^1 J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})\, ds (\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*) \right\|_{sp}$$

$$= \Big(\int_0^1 \Big\| \underbrace{J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})}_{\text{=row-stochastic, nonzero}} \Big\|_{sp\to sp} ds \Big) \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_{sp}$$

$$\leq \rho(\text{range of potentials})\, \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_{sp}$$

$\leq \tanh(\Delta(J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s}))/2) < 1$ (Birkoff's contraction Thm.)

$$\Delta(P) := \max_{i,i',j,j'} \log \frac{P_{ij}\, P_{i'j'}}{P_{ij'}\, P_{i'j}}$$

- Quick math for the convergence of Sinkhorn's Algorithm
  (Dual interpretation of Franklin and Lorenz '89)

Step 2: span-contraction of the potentials

$$\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|_{sp} = \|\xi(\boldsymbol{\beta}_{k-1}) - \xi(\boldsymbol{\beta}^*)\|_{sp} \qquad (\,sp(x) := \max x - \min x\,)$$

$$= \left\| \int_0^1 J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})\, ds (\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*) \right\|_{sp}$$

$$= \Big(\int_0^1 \Big\| \underbrace{J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s})}_{\text{=row-stochastic, nonzero}} \Big\|_{sp\to sp} ds \Big) \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_{sp}$$

$\leq \tanh(\Delta(J_{\xi;\boldsymbol{\beta}}(\boldsymbol{\beta}_{k-1;s}))/2) < 1$ (Birkoff's contraction Thm.)

$$\leq \rho(\text{range of potentials})\, \|\boldsymbol{\beta}_{k-1} - \boldsymbol{\beta}^*)\|_{sp} \qquad \Delta(P) := \max_{i,i',j,j'} \log \frac{P_{ij} P_{i'j'}}{P_{ij'} P_{i'j}}$$

$$\leq \cdots \leq \rho(\text{range of potentials})^k \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^*\|_{sp}$$

EOT on Graphs

- EOT on large graph G=(V,E)

$G = (V, E)$ a large graph, μ, ν two probability distributions on the node set V.

Find the most efficient coupling $\boldsymbol{\pi}^*$ in terms of the shortest-path distance on G

$$\pi^* = \underset{\pi \in \Pi(\mu,\nu)}{\arg\min} \langle \pi, d_G \rangle + \varepsilon D_{\mathrm{KL}}(\pi \| \mu \otimes \nu)$$

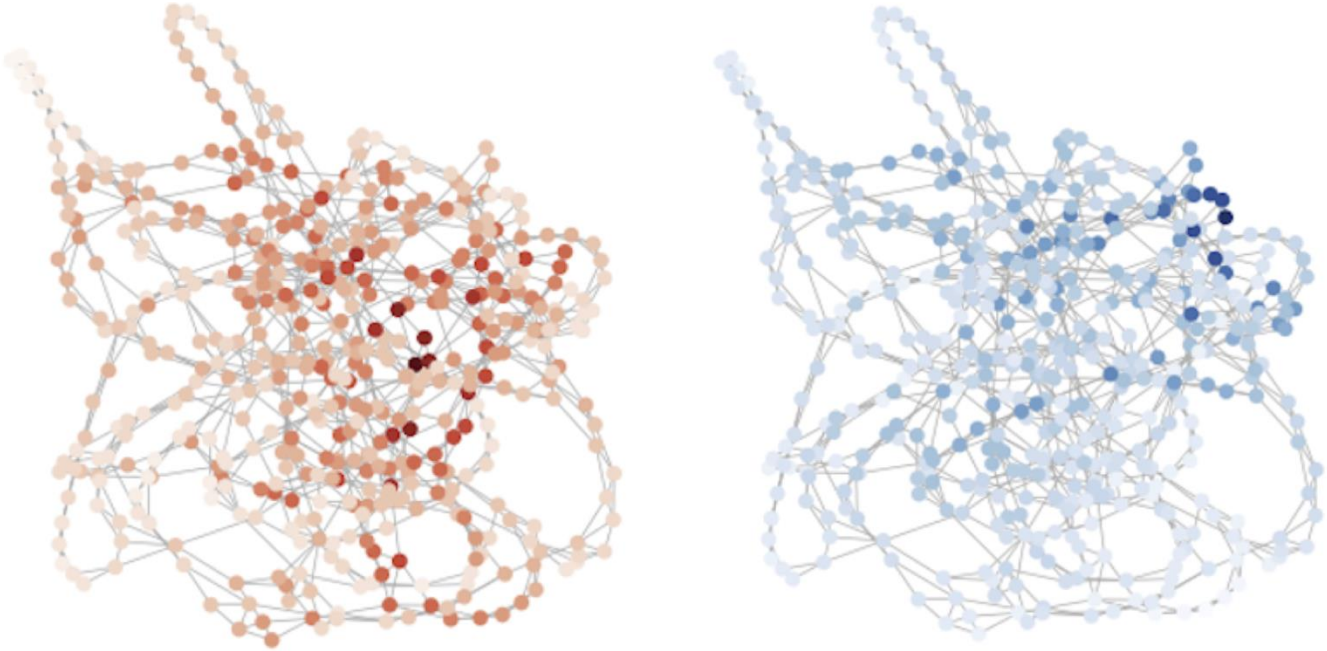


FIGURE 1. A 500-node graph G and two probability distributions μ and ν on the node set shown. Color shade indicates mass at each node.

- EOT on large graph G=(V,E)

$G = (V, E)$ a large graph, μ, ν two probability distributions on the node set V.

Find the most efficient coupling $\boldsymbol{\pi}^*$ in terms of the shortest-path distance on G

$$\pi^* = \underset{\pi \in \Pi(\mu,\nu)}{\arg\min} \langle \pi, d_G \rangle + \varepsilon D_{\mathrm{KL}}(\pi \| \mu \otimes \nu)$$

Shortest-path distance matrix

For large $G = (V, E)$, even evaluating the objective is challenging

Computational Cost: The cost matrix d_G requires an All-Pairs Shortest Path (APSP) computation = $O(n^3)$ or $O(n^2 \log n + n|E|)$

Memory Cost: The matrix d_G is dense, requiring $O(n^2)$ storage, which is prohibitive for graphs with millions of nodes.

- EOT on large graph G=(V,E)

$$(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \in \operatorname*{argmax}_{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}} f_G(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \langle \boldsymbol{\alpha}, \mu \rangle + \langle \boldsymbol{\beta}, \nu \rangle - \langle \exp(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}), K_G \odot (\mu \otimes \nu) \rangle$$

$$K_G(x, y) := \exp(-d_G(x, y)/\varepsilon)$$

- EOT on large graph G=(V,E)

$$(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \in \underset{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}}{\operatorname{argmax}} f_G(\boldsymbol{\alpha},\boldsymbol{\beta}) := \langle \boldsymbol{\alpha}, \mu\rangle + \langle \boldsymbol{\beta}, \nu\rangle - \langle \exp(\boldsymbol{\alpha}\oplus\boldsymbol{\beta}), K_G \odot (\mu\otimes\nu)\rangle$$

$$K_G(x,y) := \exp(-d_G(x,y)/\varepsilon)$$

Main Idea: Solve an easier stochastic surrogate

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \underset{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}}{\operatorname{argmax}} \left(\hat{f}_G(\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathbb{E}_{H\sim\pi}\left[\hat{f}_H(\boldsymbol{\alpha},\boldsymbol{\beta})\right]\right)$$

$$= \langle \boldsymbol{\alpha}, \mu\rangle + \langle \boldsymbol{\beta}, \nu\rangle - \langle \exp(\boldsymbol{\alpha}\oplus\boldsymbol{\beta}), \hat{K} \odot (\mu\otimes\nu)\rangle$$

where $\hat{f}_H$ = local dual objective on a random subgraph H:

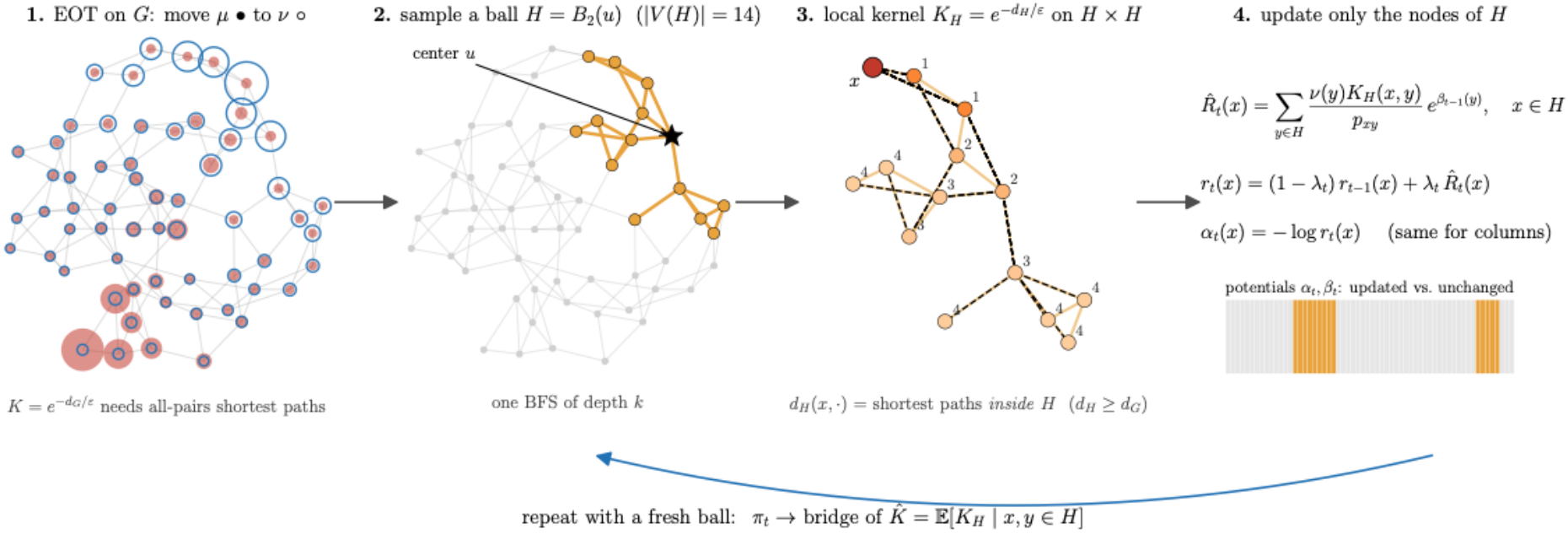
- EOT on large graph G=(V,E)

Main Idea: Solve an easier stochastic surrogate

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \operatorname*{argmax}_{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}} \left(\hat{f}_G(\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathbb{E}_{H\sim\pi}\left[\hat{f}_H(\boldsymbol{\alpha},\boldsymbol{\beta})\right]\right)$$

$$= \langle \boldsymbol{\alpha}, \mu\rangle + \langle \boldsymbol{\beta}, \nu\rangle - \langle \exp(\boldsymbol{\alpha}\oplus\boldsymbol{\beta}), \hat{K} \oplus (\mu\otimes\nu)\rangle$$

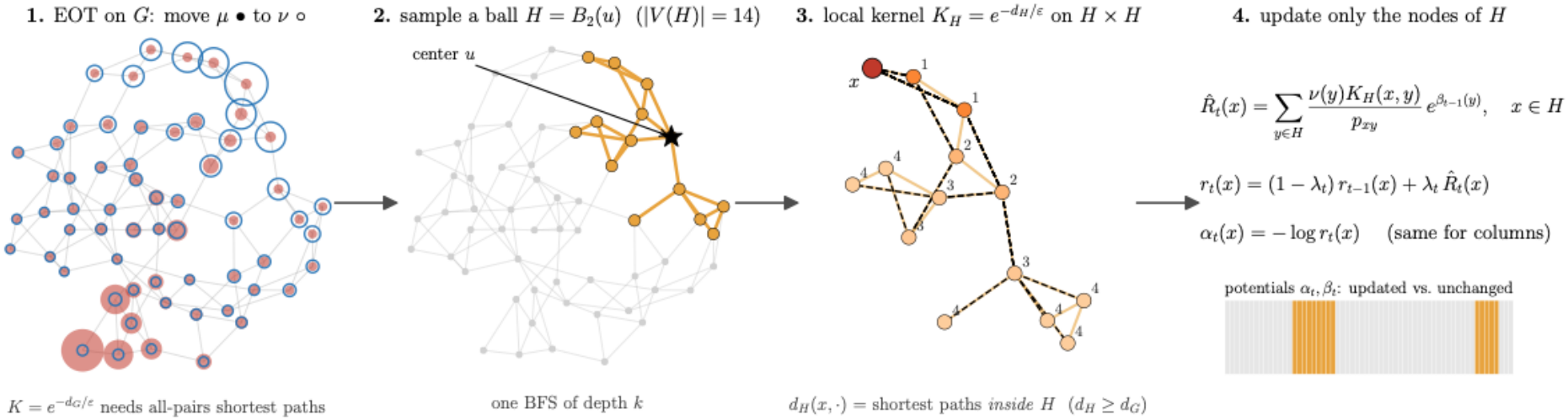
where $\hat{f}_H$ = local dual objective on a random subgraph H:



- EOT on large graph G=(V,E)

Main Idea: Solve an easier stochastic surrogate

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \operatorname*{argmax}_{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}} \left(\hat{f}_G(\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathbb{E}_{H\sim\pi}\left[\hat{f}_H(\boldsymbol{\alpha},\boldsymbol{\beta})\right]\right)$$



Q1: How do we control the stochastic relaxation bias $f_G - \hat{f}_G$?

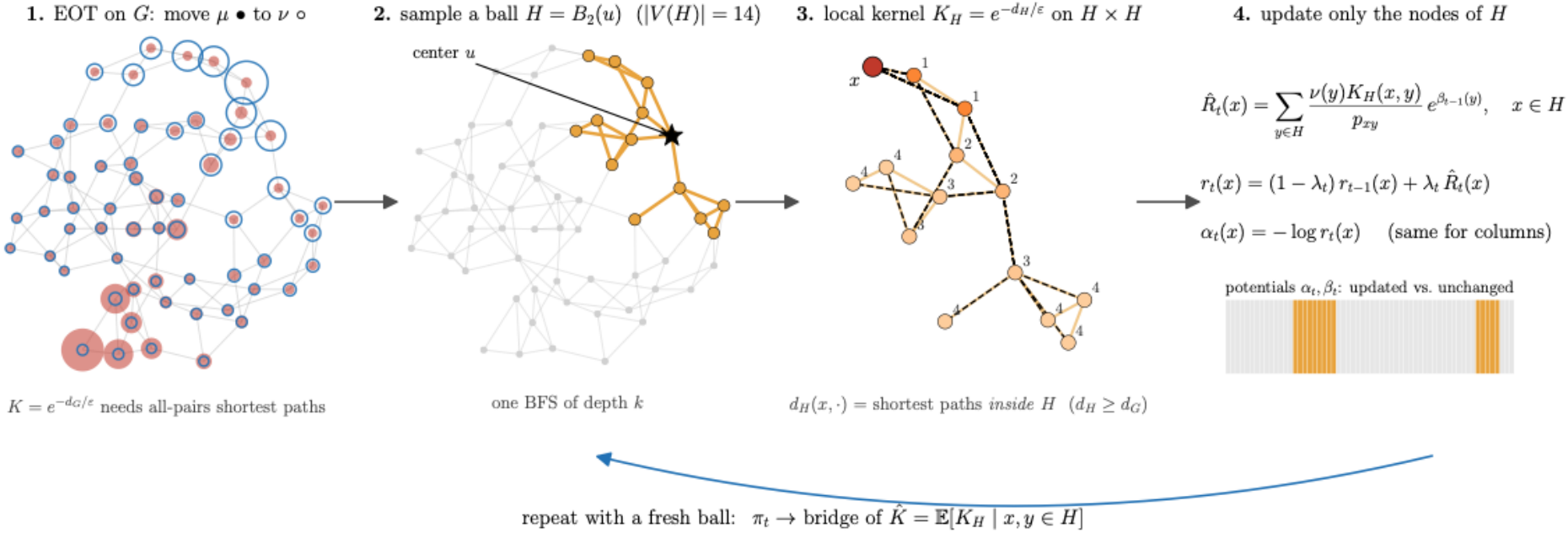
A: Stability of EOT w.r.t. kernel

Q2: How do we compute the stochastic surrogate potentials $(\hat{\alpha}, \hat{\beta})$?

A: Stochastic Sinkhorn (not today)

- EOT on large graph G=(V,E)

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \underset{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}}{\arg\max} \left(\hat{f}_G(\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathbb{E}_{H\sim\pi}\left[\hat{f}_H(\boldsymbol{\alpha},\boldsymbol{\beta}) \right] \right)$$

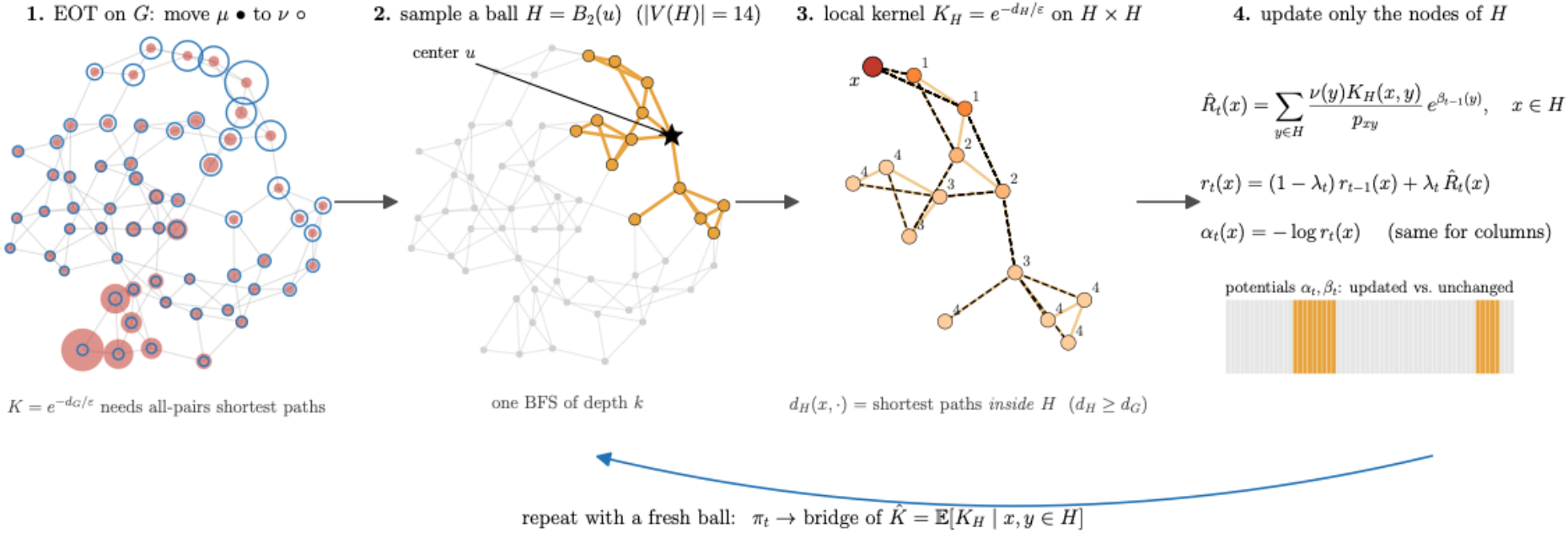


Q1: How do we control the stochastic relaxation bias $f_G - \hat{f}_G$?

A: Stability of EOT w.r.t. kernel

- EOT on large graph G=(V,E)

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \underset{\boldsymbol{\alpha},\boldsymbol{\beta}:V\to\mathbb{R}}{\arg\max} \left(\hat{f}_G(\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathbb{E}_{H\sim\pi}\left[\hat{f}_H(\boldsymbol{\alpha},\boldsymbol{\beta}) \right] \right)$$



Q1: How do we control the stochastic relaxation bias $f_G - \hat{f}_G$?

A: Stability of EOT w.r.t. kernel

Q2: How do we compute the stochastic surrogate potentials $(\hat{\alpha}, \hat{\beta})$?

A: Stochastic Sinkhorn (not today)

Q1: How do we control the stochastic relaxation bias $f_G - \hat{f}_G$?

Bias = kernel stability. Bridges are stable under kernel perturbations [Duan, Powell, L.]; for $\hat{K} \le K$, with $\eta = \log(K/\hat{K}) \ge 0$:

$$d_{\mathrm{Hell}}\left(\pi^K, \pi^{\hat{K}}\right)^2 \le \tfrac{1}{4}\,\mathbb{E}_{\pi^K}[\eta] \le \tfrac{1}{4\varepsilon}\,\mathbb{E}_{\pi^K}\left[\hat{d} - d_G\right]$$

only the *average detour* under the true bridge; no $e^{\|d_G\|_\infty/\varepsilon}$ factor.

Q1: How do we control the stochastic relaxation bias $f_G - \hat{f}_G$?

Bias = kernel stability. Bridges are stable under kernel perturbations [Duan, Powell, L.]; for $\hat{K} \le K$, with $\eta = \log(K/\hat{K}) \ge 0$:

$$d_{\mathrm{Hell}}\left(\pi^K, \pi^{\hat{K}}\right)^2 \le \tfrac{1}{4}\mathbb{E}_{\pi^K}[\eta] \le \tfrac{1}{4\varepsilon}\mathbb{E}_{\pi^K}\left[\hat{d} - d_G\right]$$

only the *average detour* under the true bridge; no $e^{\|d_G\|_\infty/\varepsilon}$ factor.

Q2: How do we compute the stochastic surrogate potentials $(\hat{\alpha}, \hat{\beta})$?

But computing $\hat{K}$ means visiting every subgraph. **Never form $\hat{K}$:** each step uses *one* sampled H, which gives an unbiased estimate of the action $\sum_y \nu(y)\hat{K}(x,y)e^{\beta(y)}$.

Q1: How do we control the stochastic relaxation bias $f_G - \hat{f}_G$?

Bias = kernel stability. Bridges are stable under kernel perturbations [Duan, Powell, L.]; for $\hat{K} \le K$, with $\eta = \log(K/\hat{K}) \ge 0$:

$$d_{\mathrm{Hell}}\big(\pi^K, \pi^{\hat{K}}\big)^2 \;\le\; \tfrac{1}{4}\,\mathbb{E}_{\pi^K}[\eta] \;\le\; \tfrac{1}{4\varepsilon}\,\mathbb{E}_{\pi^K}\big[\hat{d} - d_G\big]$$

only the *average detour* under the true bridge; no $e^{\|d_G\|_\infty/\varepsilon}$ factor.

Q2: How do we compute the stochastic surrogate potentials $(\hat{\alpha}, \hat{\beta})$?

But computing $\hat{K}$ means visiting every subgraph. **Never form $\hat{K}$:** each step uses *one* sampled H, which gives an unbiased estimate of the action $\sum_y \nu(y)\hat{K}(x,y)e^{\beta(y)}$.

Oracle: random rows I, columns J, weights W with $\mathbb{E}[W^r(x,y) \mid x \in I] = \nu(y)\bar{K}(x,y)$
($\bar{K} = \hat{K}$ for random subgraphs; also streamed samples, random features, ...)

Algorithm: average the observed actions, then take the log:

$$\hat{R}_t(x) = \textstyle\sum_y W_t^r(x,y)\, e^{\beta_{t-1}(y)}$$

$$r_t(x) = (1-\lambda_t)\, r_{t-1}(x) + \lambda_t\, \hat{R}_t(x), \qquad \alpha_t = -\log r_t$$

same for columns; visited coordinates only; gains $\lambda_t = (\#\text{visits} + 2)^{-\theta}$, $\theta \in (\tfrac{1}{2}, 1]$

Thm. (Bi, Kim, L. 26+)

If $f^{\bar{K}}$ attains its max, then a.s. (i) $\sup_t \|(\alpha_t, \beta_t)\|_\infty < \infty$; (ii) $\pi_t \to \pi^{\bar{K}}$ (exact bridge of $\bar{K}$); (iii) potentials converge up to the gauge. Hence $\limsup_t d_{\mathrm{Hell}}(\pi_t, \pi^*)^2 \le \frac{1}{4}\mathbb{E}_{\pi^*}[\eta]$.

Convergence of Stochastic Sinkhorn

Thm. (Bi, Kim, L. 26+)

If $f^{\bar{K}}$ attains its max, then a.s. (i) $\sup_t \|(\alpha_t, \beta_t)\|_\infty < \infty$; (ii) $\pi_t \to \pi^{\bar{K}}$ (exact bridge of $\bar{K}$); (iii) potentials converge up to the gauge. Hence $\limsup_t d_{\mathrm{Hell}}(\pi_t, \pi^*)^2 \le \frac{1}{4}\mathbb{E}_{\pi^*}[\eta]$.

Key technical issue : Need a priori control on the range of the iterates (α_k, β_k)

Errors $\xi = z - \hat{z}$ relative to a maximizer; $P_{xy} = \nu(y)\bar{K}(x,y)e^{\hat{\alpha}(x)+\hat{\beta}(y)}$ (and columns): $P\mathbf{1} = \mathbf{1}$, $m^\top P = m^\top$, $m = (\mu, \nu)$.

Full-batch Sinkhorn (baby case)

$$e^{-\xi_k(v)} = \sum_w P_{vw}\, e^{\xi_{k-1}(w)}$$

P stochastic $\Rightarrow$ the right side lies between $e^{\min \xi_{k-1}}$ and $e^{\max \xi_{k-1}}$:

$$\|\xi_k\|_\infty \le \|\xi_{k-1}\|_\infty$$

ℓ^∞ non-expansion, on every path
(Jacobian of the update is row-stochastic).

Stochastic Sinkhorn

$$e^{-\xi_t(v)} = (1-\lambda)\, e^{-\xi_{t-1}(v)} + \lambda \sum_w P^{\omega_t}_{vw}\, e^{\xi_{t-1}(w)}$$

P^ω is stochastic only in mean: its row sums are random (0 if H misses the row, unbounded weights) $\Rightarrow$ no pathwise bound.

Replace $\max$ by an averaged energy:

$$\mathcal{E}(z) = \sum_v m_v\, p_v^{\theta-1}\left(e^{\xi(v)} + e^{-\xi(v)}\right)$$

$\mathbb{E}[\mathcal{E}(z_t) \mid \mathcal{F}_{t-1}] \le (1+\delta_t)\,\mathcal{E}(z_{t-1})$, $\sum_t \delta_t < \infty$

Summary

Probabilistic Duality. Margin-conditioned random matrices concentrate on a static Schrödinger bridge: the primal is relative-entropy minimization, the dual is maximum likeliho od of a rank-one tilt (Kantorovich), for Poisson and general entry laws.

Scaling limit & tameness. Converging margins and prior means give a continuum Schrödinger bridge, provided the margins are tame: sharp phase transition for geometric ent ries, none for Poisson.

Optimization. Sinkhorn is alternating maximization of the dual; $\ell\infty$ non-expansion and Hilbert-metric contraction give linear convergence, at a rate set by tameness.

EOT on graphs. Replace the Gibbs kernel by expected local kernels on random k-balls; kernel stability bounds the bias, and stochastic Sinkhorn computes the surrogate.

Collaborators and Students



Sumit Mukherjee
(Columbia)



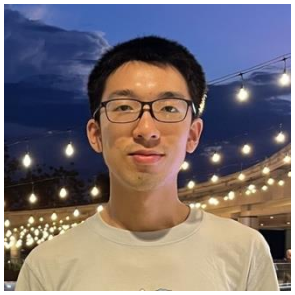
Jakwang Kim
(U. HK, Shenzhen)



William Powell
(6th year Ph.D.)



Danny Duan
(6th year Ph.D.)



Shuqi Bi
(3rd year Ph.D.)

References

Classical

G. Monge. Mémoire sur la théorie des déblais et des remblais, Hist. Acad. Sci. Paris, 1781.

L. Kantorovich. On the translocation of masses, Dokl. Akad. Nauk SSSR 37, 1942.

E. Schrödinger. Über die Umkehrung der Naturgesetze, Sitzungsber. Preuss. Akad. Wiss., 1931; Ann. Inst. H. P oincaré 2, 1932.

R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices, Ann. Math. S tatist. 35, 1964.

G. Birkhoff. Extensions of Jentzsch's theorem, Trans. AMS 85, 1957.

J. Franklin, J. Lorenz. On the scaling of multidimensional matrices, Linear Algebra Appl. 114/115, 1989.

M. Cuturi. Sinkhorn distances: lightspeed computation of optimal transport, NeurIPS 2013.

A. Barvinok. What does a random contingency table look like?, Combin. Probab. Comput. 19, 2010.

This talk

S. Dittmer, H. Lyu, I. Pak. Phase transition in random contingency tables with non-uniform margins, Trans. A MS 373, 2020.

H. Lyu, S. Mukherjee. Large random matrices with given margins, arXiv:2407.14942, 2024.

D. Duan, H. Lyu, W. Powell. Poisson contingency tables, Gibbs permutations, and entropic optimal transport , arXiv:2610.14616, 2026.

D. Duan, H. Lyu, W. Powell. Scaling limit of Sinkhorn-rescaled random matrices via stability of static Schrödi nger bridges, arXiv:2604.23461, 2026.

S. Bi, J. Kim, H. Lyu. Stochastic Sinkhorn from unbiased kernel actions, in preparation, 2026.

Thank you very much!