Average case local to global and applications

Thuy-Duong Vuong

Based on joint works with Yeganeh Alimohammadi, Nima Anari, Vishesh Jain, Yang P. Liu, Frederic Koehler, Huy Tuan Pham, Kirankumar Shiragur
Sampling from distributions

Density function $\mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, $\mathbb{P}_\mu(x) = \mu(x)/\sum_y \mu(y) \propto \mu(x)$

Approximate sampling: sample from $\hat{\mu}$ s.t.

$$d_{TV}(\mu, \hat{\mu}) = \frac{1}{2} \sum_x |\mathbb{P}_\mu(x) - \mathbb{P}_{\hat{\mu}}(x)| < 1/poly(n)$$

Counting: compute partition function $\sum_y \mu(y)$

Approximate sampling $\iff$ approximate counting
Sampling from distributions

Density function $\mu: \binom{n}{k} \rightarrow \mathbb{R}_{\geq 0}$, $P_\mu(x) = \mu(x)/\sum_y \mu(y) \propto \mu(x)$

Approximate sampling: sample from $\hat{\mu}$ s.t.
$$d_{TV}(\mu, \hat{\mu}) = \frac{1}{2} \sum_x |P_\mu(x) - P_{\hat{\mu}}(x)| < 1/poly(n)$$

Counting: compute partition function $\sum_y \mu(y)$

Approximate sampling $\iff$ approximate counting

Spin system: $Q^n \equiv \binom{[n] \times Q}{n}$
Overview

1. Background
   • Functional inequality and mixing time
   • Local to global: an inductive approach to prove functional inequality
   • Building blocks: entropic/spectral independence

2. Average-case local to global
   • Definition
   • Application: p-spin system
   • Application: spanning trees and strongly Rayleigh distributions
Multi-steps down-up walk for $\mu: \begin{pmatrix} n \\ k \end{pmatrix} \to \mathbb{R}_{\geq 0}$

1. $D_{k\to (k-\ell)}$:

2. $U_{(k-\ell)\to k}$:

$$P = D_{k\to (k-\ell)}U_{(k-\ell)\to k}$$

Glauber dynamics: $P = D_{n\to (n-1)}U_{(n-1)\to n}$
Variance/entropy contraction vs. mixing time

Variance contraction
• $\chi^2(\nu P || \mu P) \leq (1 - \rho_{\chi^2})\chi^2(\nu || \mu)$

• $\rho_{\chi^2} = 1 - \lambda_2(P)$

Entropy contraction
• $\mathcal{D}_{KL}(\nu P || \mu P) \leq (1 - \rho_{KL})\mathcal{D}_{KL}(\nu || \mu)$

$P = D_{k \rightarrow (k-\ell)} U_{(k-\ell) \rightarrow k}$
Variance/entropy contraction vs. mixing time

**Variance contraction**
- $\chi^2(vP||\mu P) \leq (1 - \rho_{\chi^2})\chi^2(v||\mu)$
- $T_{mix} \leq \rho_{\chi^2}^{-1} \log (\min \mu(x))^{-1} \approx \rho_{\chi^2}^{-1} n$
- $\rho_{\chi^2} = 1 - \lambda_2(P)$

**Entropy contraction**
- $\mathcal{D}_{KL}(vP||\mu P) \leq (1 - \rho_{KL})\mathcal{D}_{KL}(v||\mu)$
- $T_{mix} \leq \rho_{KL}^{-1} \log \log (\min \mu(x))^{-1} \approx \rho_{KL}^{-1} \log n$

Entropy contraction implies variance contraction: $\rho_{\chi^2} \geq \rho_{KL} \geq 0$

Typically for Glauber dynamics: $\rho_{\chi^2} \approx \rho_{KL} = \frac{1}{n}$

but $\log \min \mu(x)^{-1} \approx n$. Bounding $\rho_{KL} \Rightarrow$ quadratic improvement on $T_{mix}$

It is hard to bound $\rho_{KL}$!
Overview

1. Background
   • Functional inequality and mixing time
   • Local to global: an inductive approach to prove functional inequality
   • Building blocks: entropic/spectral independence

2. Average-case local to global
   • Definition
   • Application: p-spin system
   • Application: spanning trees and strongly Rayleigh distributions
Local-to-global

$$\mu: \binom{[n]}{k} \rightarrow \mathbb{R}_{\geq 0}$$

$\mu$ is reducible i.e. conditionals $\mu(. | S): \binom{[n]}{k-|S|} \rightarrow \mathbb{R}_{\geq 0}$ of $\mu$ has same property as $\mu$

$$\mu(S'|S) \propto \mu(S' \cup S)$$

Local-to-global (variation of [Alev-Lau—STOC’21])

Entropy contraction

$$\forall \nu: D_{\text{KL}}(\nu||\mu) \geq (1 + \gamma_k)D_{\text{KL}}(\nu D_{k \rightarrow (k-\ell)} || \mu D_{k \rightarrow (k-\ell)})$$
Local-to-global

$$\mu: \binom{n}{k} \rightarrow \mathbb{R}_{\geq 0}$$

Local-to-global (variation of [Alev-Lau—STOC’21])

Entropy contraction

$$\forall \nu: \mathcal{D}_{KL}(\nu||\mu) \geq (1 + \gamma_k)\mathcal{D}_{KL}(\nu D_{k\rightarrow(k-\ell)}||\mu D_{k\rightarrow(k-\ell)})$$

Inductive hypothesis on $$\mu(|i|): \binom{n}{k-1} \rightarrow \mathbb{R}_{\geq 0}, \mu(S|i) \propto \mu(S)$$

$$\mathcal{D}_{KL}(\nu(\cdot|i)||\mu(\cdot|i)) \geq (1 + \gamma_{k-1})\mathcal{D}_{KL}(\nu(\cdot|i) D_{(k-1)\rightarrow(k-\ell-1)}||\mu D_{(k-1)\rightarrow(k-\ell-1)})$$
Local-to-global

\( \mu: \binom{n}{k} \rightarrow \mathbb{R}_{\geq 0} \)

\( D_{k \rightarrow 1}(S) \): sample \( i \in S \) uniformly

\( \nu D_{k \rightarrow 1}(i) = \text{marginal of } i \propto \mathbb{P}_{S \sim \nu}[i \in S] = \nu(i) \)

Local-to-global

Entropy contraction

Inductive hypothesis on \( \mu(. | i): \binom{n}{k-1} \rightarrow \mathbb{R}_{\geq 0}, \mu(S|i) \propto \mu(S) \)

\[ \mathcal{D}_{K-L}(\nu || \mu) \geq (1 + \gamma_{k-1}) \mathcal{D}_{K-L}(\nu D_{k \rightarrow (k-1)} || \mu D_{k \rightarrow (k-1)}) \]

\[ \mathcal{D}_{K-L}(\nu(i) || \mu(i)) \geq (1 + \gamma_{k-1}) \mathcal{D}_{K-L}(\nu(i) D_{(k-1) \rightarrow (k-1)} || \mu D_{(k-1) \rightarrow (k-1)}) \]

\[ \sum_i \nu(i)(\mathcal{D}_{K-L}(\nu(i) || \mu(i)) - (1 + \gamma_{k-1}) \mathcal{D}_{K-L}(\nu(i) D_{(k-1) \rightarrow (k-1)} || \mu D_{(k-1) \rightarrow (k-1)})) \geq 0 \]
Local-to-global

\[ \mu: \left( \begin{bmatrix} n \\ k \end{bmatrix} \right) \rightarrow \mathbb{R}_{\geq 0} \]

\( D_{k \rightarrow 1}(S) \): sample \( i \in S \) uniformly

\( \frac{1}{\alpha} \)-entropic independence \( \iff \forall \nu: D_{KL}(\nu || \mu) \geq \alpha k D_{KL}(\nu D_{k \rightarrow 1} || \mu D_{k \rightarrow 1}) \)

\( \nu D_{k \rightarrow 1}(i) = \text{marginal of } i \propto \mathbb{P}_{S \sim \nu}[i \in S] = \nu(i) \)

Local-to-global

Entropy contraction

Inductive hypothesis on \( \mu(, |i): \left( \begin{bmatrix} n \setminus \{i\} \\ k-1 \end{bmatrix} \right) \rightarrow \mathbb{R}_{\geq 0}, \mu(S|i) \propto \mu(S) \)

\[ \sum \nu(\cdot|\cdot) D_{KL}(\nu(\cdot|\cdot) || \mu(\cdot|\cdot)) \geq (1 + \gamma_{k-1}) D_{KL}(\nu(\cdot|\cdot) D_{k \rightarrow (k-\ell-1)-1} || \mu D_{k \rightarrow (k-\ell-1)}) \]

\[ D_{KL}(\nu || \mu) + \gamma_{k-1} D_{KL}(\nu D_{k \rightarrow 1} || \mu D_{k \rightarrow 1}) \geq (1 + \gamma_{k-1}) D_{KL}(\nu D_{k \rightarrow 1} || \mu D_{k \rightarrow 1}) \]
Overview

1. Background
   • Functional inequality and mixing time
   • Local to global: an inductive approach to prove functional inequality
   • Building blocks: entropic/spectral independence

2. Average-case local to global
   • Definition
   • Application: p-spin system
   • Application: spanning trees and strongly Rayleigh distributions
Entropic independence

\[ \mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0} \]

\( D_{k\to 1}(S) \): sample \( i \in S \) uniformly

\( \frac{1}{\alpha} \)-entropic independence \( \iff \forall \nu: D_{KL}(\nu||\mu) \geq \alpha k D_{KL}(\nu D_{k\to 1}||\mu D_{k\to 1}) \)

Thm: If \( \mu_S = \mu(.|S) \) are \( \frac{1}{\alpha} \)-entropic independence \( \forall S \)

\( \Rightarrow \) entropy contraction of \( D_{k \to (k-\ell)} \)

\( \Rightarrow \) optimal bound for modified Log-Sobolev constant and mixing time of down-up walks
Spectral independence

\[ \mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0} \]

\[ D_{k\to 1}(S): \text{sample } i \in S \text{ uniformly} \]

\( \frac{1}{\alpha} \)-spectral independence \( \iff \forall \nu: D_{x^2}(\nu||\mu) \geq \alpha k D_{x^2}(\nu D_{k\to 1}||\mu D_{k\to 1}) \)

Thm: If \( \mu_S = \mu(.,|S) \) are \( \frac{1}{\alpha} \)-spectral independence \( \forall S \)

\( \Rightarrow \) variance contraction of \( D_{k\to (k-\ell)} \)

\( \Rightarrow \) optimal bound for spectral gap/Poincare constant of down-up walks
What if $\mu(\cdot|S)$ is not entropically independent for some $S$?
Overview

1. Background
   • Functional inequality and mixing time
   • Local to global: an inductive approach to prove functional inequality
   • Building blocks: entropic/spectral independence

2. Average-case local to global
   • Definition
   • Application: pspin
   • Application: spanning trees and strongly Rayleigh distributions
Average local-to-global

Thm [Alimohammadi-Anari-Shiragur-V’21, Anari-Liu-V’22]: If $\mu_T = \mu(\cdot|T)$ are $(k - |T|)(1 - \rho(T))$-spectral independence $\forall T$

$$(1 - \rho(T))D_{KL}(\nu||\mu_T) \geq D_{KL}(\nu D_{k \rightarrow 1}||\mu_T D_{k \rightarrow 1})$$

Harmonic mean:

$$\gamma_T = \mathbb{E}_{e_1, \ldots, e_{|T|} \sim \text{permutation}(T)} \left[ (\rho(\emptyset)\rho(\{e_1\}) \ldots \rho(\{e_1, \ldots, e_{|T|-1}\}))^{-1} \right]^{-1}$$

Then for $\kappa = \min \{\gamma_T | T \in \binom{[n]}{\ell} \}$

$$(1 - \kappa)D_{KL}(\nu||\mu) \geq D_{KL}(\nu D_{k \rightarrow (k-\ell)}||\mu D_{k \rightarrow (k-\ell)})$$
Average local to global

Intuition:
For each fixed set $T \in \binom{[n]}{\ell}$ and $s$,
If for average $S \in \binom{T}{s}$, $\mu_S$ is entropically independent
then $D_{k \rightarrow (k-\ell)}$ has good entropy contraction
Average local to global

Intuition:

For each fixed set $T \in \binom{\mathcal{X}}{\ell}$ and $s$,

If for average $S \in \binom{T}{s}$, $\mu_S$ is entropically independent
then $D_{k \rightarrow (k-\ell)}$ has good entropy contraction?
Not true!
Overview

1. Background
   • Functional inequality and mixing time
   • Local to global: an inductive approach to prove functional inequality
   • Building blocks: entropic/spectral independence

2. Average-case local to global
   • Definition
   • Application: p-spin
   • Application: spanning trees and strongly Rayleigh distributions
Spin system

$n$ vertices, each assigned spin $+1$ or $-1$.

Distribution over configurations $\sigma \in \{-1,1\}^n$

Density function:

$$\mu: \{-1,1\}^n \rightarrow \mathbb{R}_{\geq 0}$$

Gibbs measure: $\mu(\sigma) = \exp(H(\sigma))$
Spin system

$n$ vertices, each assigned spin $+1$ or $-1$.

Density function:

$$\mu: \{-1, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$$

Ising model: pairwise interaction (graph edges)

$$H(\sigma) = \frac{\beta}{\sqrt{n}} \sum_{i \leq j} J_{ij} \sigma_i \sigma_j + \sum_i h_i \sigma_i$$

P-spin: interaction between $p$ vertices (hypergraph edges)

$$H(\sigma) = \sum_{p,i_1,\ldots,i_p} \frac{\beta_p}{(p-1)!} J_{i_1,\ldots,i_p} \sigma_1 \ldots \sigma_p + \sum_i h_i \sigma_i$$
Spin system

$n$ vertices, each assigned spin $+1$ or $-1$.

Density function:

$\mu: \{-1,1\}^n \rightarrow \mathbb{R}_{\geq 0}$

SK Ising model: pairwise interaction (graph edges)

$$H(\sigma) = \frac{\beta}{\sqrt{n}} \sum_{i,j} J_{ij} \sigma_i \sigma_j + \sum_i h_i \sigma_i$$

P-spin: interaction between $p$ vertices (hypergraph edges)

$$H(\sigma) = \sum_{p,i_1,\ldots,i_p} \frac{\beta_p}{\binom{n}{p-1}^2} J_{i_1,\ldots,i_p} \sigma_1 \ldots \sigma_p + \sum_i h_i \sigma_i$$

Gaussians $\mathcal{N}(0,1)$
Glauber dynamics on p-spin

- Warmup, $p = 2$: if $0 \leq \beta \leq \epsilon$, optimal $O_\beta(n \log n)$ mixing of Glauber dynamics
- If $\sum_p \sqrt{p^3 \log p \beta_p} \leq \epsilon$. Let $\beta = \sum_p \sqrt{2p^3 \log p \beta_p}$, optimal $O_\beta(n \log n)$ mixing of Glauber dynamics [Anari-Jain-Koehler-Pham-V’23]
  - [Adhikari, Brennecke, Xu, Yau’22] spectral gap $\geq \Omega_\beta \left(\frac{1}{n}\right) \Rightarrow O_\beta(n^2 \log n)$ mixing
Thm [Anari-Jain-Koehler-Pham-V’23]: \( \epsilon = \theta(1). \mu(\sigma) = \exp(H(\sigma)) \).

If \( \beta = \max_{\sigma \in \{\pm 1\}^n} \| \nabla^2 H(\sigma) \|_{op} \leq \epsilon \) then

\( \mu^{\text{hom}} \) is \((1 + O(\beta))\)-entropically independent.
Thm [Anari-Jain-Koehler-Pham-V'23]: \( \epsilon = \Theta(1) \). \( \mu(\sigma) = \exp(H(\sigma)) \).

If \( \beta = \max_{\sigma \in \{\pm 1\}^n} \| \nabla^2 H(\sigma) \|_{op} \leq \epsilon \) then \( \mu^{hom} \) is \((1 + O(\beta))\)-entropically independent

Conditionals of \( \mu^{hom} \equiv \) pinnings of \( \mu \): \( \mu(. | \sigma_S = \tau) \)

\( \mu^{hom}(. | \sigma_S = \tau) \) is \((1 + O(\beta))\)-entropically independent
Thm [Anari-Jain-Koehler-Pham-V’23]: $\epsilon = \theta(1). \mu(\sigma) = \exp(H(\sigma))$.

If $\beta = \max_{\sigma \in \{\pm 1\}^n} \|\nabla^2 H(\sigma)\|_{op} \leq \epsilon$ then $\mu^{\text{hom}}$ is $(1 + O(\beta))$-entropically independent.

Conditionals of $\mu^{\text{hom}} \equiv$ pinnings of $\mu$: $\mu(. | \sigma_S = \tau) = \exp(H_{S,\tau}(\sigma))$

$$\beta_{S,\tau} = \max_{\sigma_{Sc} \in \{\pm 1\}^n} \|\nabla^2 H_{S,\tau}(\sigma_{Sc})\|_{op} \leq \beta$$

$\mu^{\text{hom}}(. | \sigma_S = \tau)$ is $(1 + O(\beta))$-entropically independent

$\Rightarrow \tilde{O}(n^{1+O(\beta)})$ mixing time
Thm [Anari-Jain-Koehler-Pham-V’23]: \( \epsilon = \theta(1). \mu(\sigma) = \exp(H(\sigma)). \)

If \( \beta = \max_{\sigma \in \{\pm 1\}^n} \left\| \nabla^2 H(\sigma) \right\|_{op} \leq \epsilon \) then \( \mu^{\text{hom}} \) is \( (1 + O(\beta)) \)-entropically independent

Conditionals of \( \mu^{\text{hom}} \equiv \) pinnings of \( \mu: \mu(. \mid \sigma_S = \tau) = \exp(H_{S,\tau}(\sigma)) \)

\[
\beta_{S,\tau} = \max_{\sigma_S \in \{\pm 1\}^n} \left\| \nabla^2 H_{S,\tau}(\sigma_{SC}) \right\|_{op} \leq \beta
\]

\( p = 2: H_{S,\tau}(\sigma) = \frac{\beta}{\sqrt{n}} \sum_{i,j \in SC} J_{ij} \sigma_i \sigma_j + \sum_i h' \sigma_i \)

\[
\beta_{S,\tau} \leq \beta \sqrt{\frac{|SC|}{n}} \Rightarrow \tilde{O}(n) \text{ mixing time}
\]
Conditionals of $\mu^{\text{hom}} \equiv$ pinnings of $\mu$: $\mu(. | \sigma_S = \tau) = \exp(H_{S,\tau}(\sigma))$

$$\beta_{S,\tau} = \max_{\sigma_{Sc} \in \{\pm 1\}^n} \left| \nabla^2 H_{S,\tau}(\sigma_{Sc}) \right|_{op} \leq \beta$$

$p = 3$: $H_{S,\tau}(\sigma) = \frac{\beta_3}{n} \left( \sum_{i,j,k \in S^c} J_{ijk} \sigma_i \sigma_j + \sum_{i,j \in S^c, k \in S} J_{ijk} \sigma_i \sigma_j \tau_k \right) + \sum_i h' i \sigma_i$

Hope: $\beta_{S,\tau} \leq \beta \sqrt{\frac{|S^c|}{n}}$. But, there are bad pinnings! E.g. $S^c = \{1,2\}$

$H_{S,\tau}(\sigma) = \frac{\beta_3}{n} \sum_{k \in S} J_{12k} \tau_k \sigma_i \sigma_j$
Conditionals of $\mu^{\text{hom}} \equiv$ pinnings of $\mu$: 
$$\mu(\cdot|\sigma_S = \tau) = \exp(H_{S,\tau}(\sigma))$$

$$\beta_{S,\tau} = \max_{\sigma_{Sc} \in \{\pm 1\}^n} \left| \nabla^2 H_{S,\tau}(\sigma_{Sc}) \right|_{op} \leq \beta$$

$p = 3$: 
$$H_{S,\tau}(\sigma) = \frac{\beta_3}{n} \left( \sum_{i,j,k \in S^c} J_{ijk} \sigma_i \sigma_j + \sum_{i,j \in S^c, k \in S} J_{ijk} \sigma_i \sigma_j \tau_k \right) + \sum_i h'_i \sigma_i$$

Hope: $\beta_{S,\tau} \leq \beta \sqrt{\frac{|S^c|}{n}}$. But, there are bad pinnings! E.g. $S^c = \{1,2\}$

$$H_{S,(\text{sign}(J_{12k}))_k}(\sigma) = \frac{\beta_3}{n} \sum_{k \in S} |J_{12k}| \sigma_i \sigma_j \text{ thus } \beta_{S,\tau} \approx \beta_3 = \theta(1)$$
Conditionals of $\mu_{h} \equiv \text{pinnings of } \mu: \mu(. | \sigma_S = \tau) = \exp(H_{S,\tau}(\sigma))$

$$\beta_{S,\tau} = \max_{\sigma_{S^c} \in \pm 1} \left\| \nabla^2 H_{S,\tau}(\sigma_{S^c}) \right\|_{op} \leq \beta$$

$p = 3$: $H_{S,\tau}(\sigma) = \frac{\beta_3}{n} (\sum_{i,j,k \in S^c} J_{ijk} \sigma_i \sigma_j + \sum_{i,j \in S^c, k \in S} J_{ijk} \sigma_i \sigma_j \tau_k) + \sum_i h'_i \sigma_i$

Hope: $\beta_{S,\tau} \leq \beta \sqrt{\frac{|S^c|}{n}}$. But, there are bad pinnings! E.g. $S^c = \{1,2\}$

$$H_{S, (\text{sign}(J_{12k}))_k}(\sigma) = \frac{\beta_3}{n} \sum_{k \in S} |J_{12k}| \sigma_i \sigma_j \text{ thus } \beta_{S,\tau} \approx \beta_3 = \theta(1)$$

But, “most” pinnings are good!
Overview

1. Background
   • Functional inequality and mixing time
   • Local to global: an inductive approach to prove functional inequality
   • Building blocks: entropic/spectral independence

2. Average-case local to global
   • Definition
   • Application: p-spin system
   • Application: spanning trees and strongly Rayleigh distributions
Sampling random spanning trees

Given $G$, output spanning tree $T$ with probability $\frac{1}{\#\text{spanning-trees}}$.

To find one spanning tree, need $\Omega(|E|)$ time.

$\Rightarrow$ need $\Omega(|E|)$ time to sample.

[Anari-Liu-OveisGharan-Vinzant-Vuong'21] $O(|E|\log^2|E|)$ using up-down walk.
Sampling random spanning trees

Given G, output spanning tree $T$ with probability $\frac{1}{\#\text{spanning-trees}}$

Can we produce sample in sublinear time after preprocessing?
Sampling random spanning trees

Given G, output spanning tree $T$ with probability $\frac{1}{\#\text{spanning-trees}}$

Can we produce sample in sublinear time after preprocessing?

[Anari-Liu-V--FOCS'22]: sample in $O(|V| \log^2 |V|)$ time after $O(|E| \log^2 |V|)$ preprocessing
Up-down walk

Repeat for sufficiently many times. Take tree $T$

1. Add an edge $e$
2. Remove an edge $f$ uniformly at random from the unique circle in $T + e$
Up-down walk

Repeat for sufficiently many times. Take tree T
1. Add an edge e
2. Remove an edge f uniformly at random from the unique circle in $T + e$

Up-down walk $\equiv$ down-up walk on the complement defined by
$$\bar{\mu}([n] \setminus S) = \mu(S)$$
Up-down walk

Repeat for sufficiently many times. Take tree $T$

1. Add an edge $e$

2. Remove an edge $f$ uniformly at random from the unique circle in $T + e$. Update $T \leftarrow T + e - f$

Key points:

- Can implement 1 and 2 in $O(\log |V|)$-time using link-cut tree
- If $\exists$ bridge edge, need $\theta(|E| \log |E|)$ time to converge
Up-down walk

Repeat for sufficiently many times. Take tree $T$

1. Add an edge $e$

2. Remove an edge $f$ uniformly at random from the unique circle in $T + e$. Update $T \leftarrow T + e - f$

Key points:
• Can implement 1 and 2 in $O(\log |V|)$-time using link-cut tree
• If $\exists$ bridge, need $\theta(|E| \log |E|)$ time to mix
• If all edges have same marginal, mixes in $O(|V| \log |V|)$ time
Isotropic transformation

Goal: make all edges/elements having the same marginal.

\[ \mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0} \]

Let \( p_e = Pr_\mu [e \in T] \). Replace edge \( e \) with \( t_e = \lceil \frac{np_e}{k} \rceil \) parallel edges \( e' \).
Strongly Rayleigh distributions

\( D_{k \to 1}(S) \): sample \( i \in S \) uniformly

\( \mu \) is \( \frac{1}{\alpha} \)-entropic independence \( \iff \forall \nu: \)

\[
\mathcal{D}_{KL}(\nu || \mu) \geq \alpha k \mathcal{D}_{KL}(\nu D_{k \to 1} || \mu D_{k \to 1})
\]

\( \mu \) strongly Rayleigh \( \Rightarrow \) 1-entropic independence

\[
\mathcal{D}_{KL}(\nu || \mu) \geq k \mathcal{D}_{KL}(\nu D_{k \to 1} || \mu D_{k \to 1})
\]

Examples:

- \( U(\{\text{spanning trees}\}) \)
- Determinantal point processes:
Improved entropic independence under uniform marginals

\( \mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0} \) strongly Rayleigh. When \( p_e \leq \tilde{O} \left( \frac{k}{n} \right) \) \( \forall e \in [n] \)

\[ \mathcal{D}_{KL} (\bar{\nu} || \bar{\mu}) \geq (n - k) \log(n/k) \mathcal{D}_{KL} (\nu_{D_{(n-k)\to 1}} || \mu_{D_{(n-k)\to 1}}) \]
Improved entropic independence under uniform marginals

\( \mu: \binom{[n]}{k} \rightarrow \mathbb{R}_{\geq 0} \) strongly Rayleigh. When \( p_e \leq \tilde{O} \left( \frac{k}{n} \right) \forall e \in [n] \)

\[
D_{KL}(\tilde{\nu}||\tilde{\mu}) \geq (n - k) \log \left( \frac{n}{k} \right) D_{KL}(\tilde{\nu}D_{(n-k)\rightarrow 1}||\tilde{\mu}D_{(n-k)\rightarrow 1})
\]

1. \( D_{KL}(\tilde{\nu}||\tilde{\mu}) = D_{KL}(\nu||\mu) \geq kD_{KL}(\nuD_{k\rightarrow 1}||\muD_{k\rightarrow 1}) \)
Improved entropic independence under uniform marginals

\( \mu: \binom{[n]}{k} \rightarrow \mathbb{R}_{\geq 0} \) strongly Rayleigh. When \( p_e \leq \tilde{O}\left(\frac{k}{n}\right) \forall e \in [n] \)

\[
D_{KL}(\tilde{\nu}||\tilde{\mu}) \geq (n - k) \log \left(\frac{n}{k}\right) D_{KL}(\tilde{\nu} D_{(n-k)\rightarrow 1}||\tilde{\mu} D_{(n-k)\rightarrow 1})
\]

1. \( D_{KL}(\tilde{\nu}||\tilde{\mu}) = D_{KL}(\nu||\mu) \geq k D_{KL}(\nu D_{k\rightarrow 1}||\mu D_{k\rightarrow 1}) \)

2. \( k D_{KL}(\nu D_{k\rightarrow 1}||\mu D_{k\rightarrow 1}) \geq (n - k) \log \left(\frac{n}{k}\right) D_{KL}(\tilde{\nu} D_{(n-k)\rightarrow 1}||\tilde{\mu} D_{(n-k)\rightarrow 1}) \)

Here we use the uniform marginal assumption.
Improved EI implies improved mixing time

Entropy contraction of $D_{(n-k) \to 1}$ for $\bar{\mu}$ and its conditionals

$\Rightarrow$ Entropy contraction of $D_{(n-k) \to (n-k-1)}$

$\Rightarrow$ Mixing time of up-down walk.

$(n - k)$ contraction $\Rightarrow n \log n$ mixing time ☹

$(n - k) \log\left(\frac{n}{k}\right)$ contraction $\Rightarrow k \log n$ mixing time 😊
But, not all conditionals of $\bar{\mu}$ has improved entropy contraction \( \otimes \) i.e. exists $\bar{S}$ s.t.

$$
D_{KL}(\bar{\nu}_{\bar{S}}||\bar{\mu}_{\bar{S}}) < (n - k) \log \left( \frac{n}{k} \right) D_{KL}(\bar{\nu}_{\bar{S}}D_{(n-k)\rightarrow 1}||\bar{\mu}_{\bar{S}}D_{(n-k)\rightarrow 1})
$$

\[ \bar{S} = \{ \bar{1}, \bar{b} \} \]
Average local to global

For each set $\bar{W} \in \binom{[n]}{n-k-1}$ and $s$, if for “many” $\bar{S} \in \binom{\bar{W}}{n-s}$ $\bar{\mu}_{\bar{S}}$ has uniform marginal thus improved entropy contraction then we still get $k \log n$ mixing time 😊

“many” = w/ prob. $1 - 1/n^{10}$ over uniformly chosen $\bar{S}$
Average local to global

For each set $\tilde{W} \in \binom{[n]}{n-k-1}$ and $s$, if for “many” $\tilde{S} \in \binom{\tilde{W}}{n-s}$
$\tilde{\mu}_\tilde{S}$ has uniform marginal thus improved entropy contraction

Proof:

Compare marginals of $\tilde{\mu}_\tilde{S}$ and $\tilde{\mu}_{\tilde{S} \cup \{s'\}}$ for random $s'$

Since $\mu$ is strongly Rayleigh, marginal doesn’t change much

Use martingale argument and Bernstein ineq.