Optimal Sublinear Sampling of Spanning Trees and Determinantal Point Processes via Average-Case Entropic Independence

Thuy-Duong Vuong

Joint with Nima Anari Yang Liu

Sampling from a distribution

- Given access to density function $\mu: \Omega \to \mathbb{R}_{\geq 0}$, output x in Ω s.t.

$$\mathbb{P}[x] \propto \mu(x)$$

E.g.: (sampling problems)

- random spanning tree [Aldous-Broder'90, Colbourn-Myrvold-Neufeld'96, Kelner-Madry'09, Madry-Straszak-Tarnawski'15, Schild'18, Anari-Liu-OveisGharan-Vinzant-**V.**—STOC'21]
- matroid bases [Anari-Liu-OveisGharan-Vinzant—STOC'19,Cryan-Guo-Mousa—FOCS'19]

Sampling from a distribution

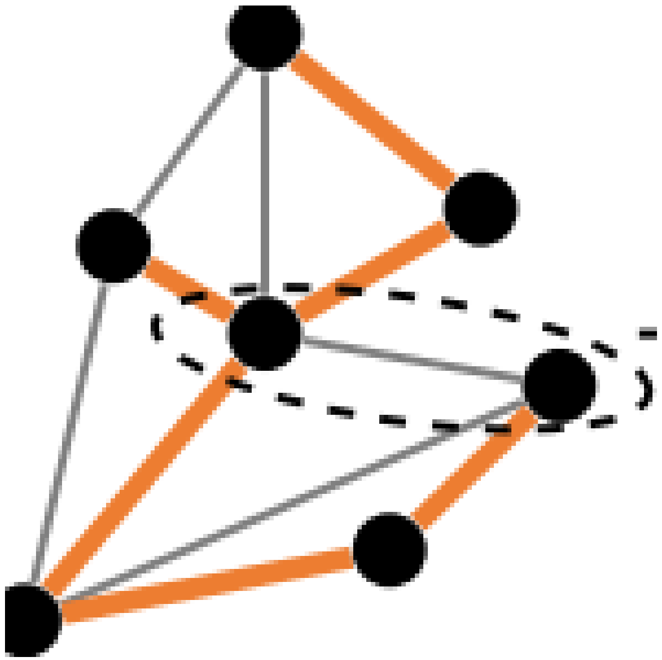
- Given access to density function $\mu: \Omega \to \mathbb{R}_{\geq 0}$, output x in Ω s.t.
$$\mathbb{P}[x] \propto \mu(x)$$
- Sufficient to approximately sample i.e. output x according to $\widehat{\mathbb{P}}$ s.t.
$$d_{TV}\left(\mathbb{P}, \widehat{\mathbb{P}}\right) = \sum |\widehat{\mathbb{P}}(x) - \mathbb{P}(\mathrm{x})| < 0.01$$

Overview

1. Motivation
   - Random spanning trees
   - Determinantal point processes
2. Algorithm
   - Isotropic transformation
   - Up-down walk
3. Analysis
   - Improved entropic independence under uniform marginals
   - Mixing time bound using average case local-to-global argument

Spanning tree

Given graph G = G(V,E), $T \subseteq E$ is a spanning tree of G if T has no loop and $|T| = |V| - 1$.



Sampling random spanning trees

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

Sampling random spanning trees

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

Application:

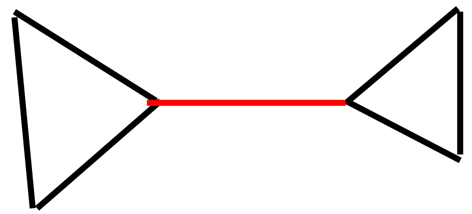
- Travelling salesman (Christofides algorithm, [Karlin-Klein-OveisGharan'21])
- Graph sparsification [Goyal-Rademacher-Vempala'09,Kyng-Song'18]

Sampling random spanning trees

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

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

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



Sampling random spanning trees

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

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

$\Rightarrow$ need $\Omega(|E|)$ time to sample. This is tight!

[Schild'18] $O(|E|^{1+\epsilon})$ for any $\epsilon > 0$.

[Anari-Liu-OveisGharan-Vinzant-Vuong'21] $O(|E|\log^2|E|)$ Simpler algorithm + analysis

Sampling random spanning trees

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

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

$\Rightarrow$ need $\Omega(|E|)$ time to sample. This is tight!

[Schild'18] $O(|E|^{1+\epsilon})$ for any $\epsilon > 0$.

[Anari-Liu-OveisGharan-Vinzant-Vuong'21] $O(|E| \log^2 |E|)$ Simpler algorithm + analysis

Can we do better than linear time?

Sampling random spanning trees

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

Can we produce sample in sublinear time after preprocessing?

Sampling random spanning trees

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

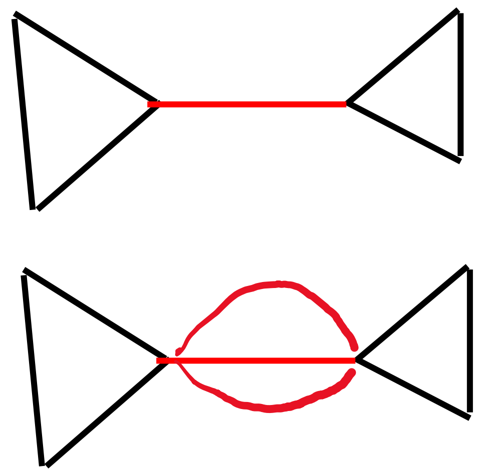
Can we produce sample in sublinear time after preprocessing?

YES!

Idea:

1. Make all edges equal (having the same marginal under the uniform spanning tree distributions) via subdividing edges

1. Apply sampling algorithm from [Anari-Liu-OveisGharan-Vinzant-Vuong'21] (up-down walk)



Sampling random spanning trees

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

Can we produce sample in sublinear time after preprocessing?

YES!

[Anari-Liu-Vuong--FOCS'22]

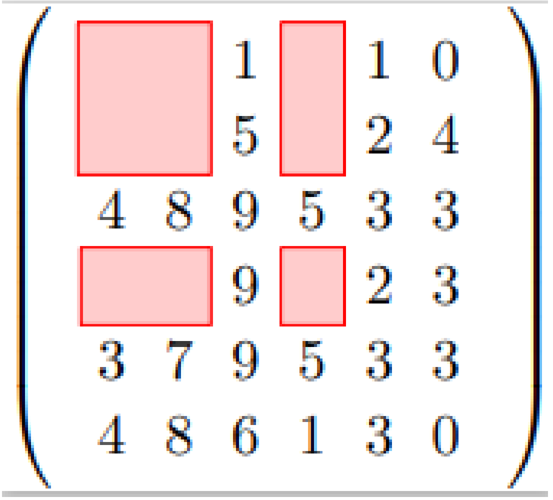
After $\tilde{O}(|E|)$ preprocessing time, can sample a random spanning tree in $\tilde{O}(|V|)$ time.

Other applications

Determinantal point processes (DPP):

(PSD) matrix $L \in \mathbb{R}^{n \times n}$: $\mu_L(S) = \det(L_{S,S})$ for $|S| = k$

Applications: random linear algebra, machine learning (recommender system)

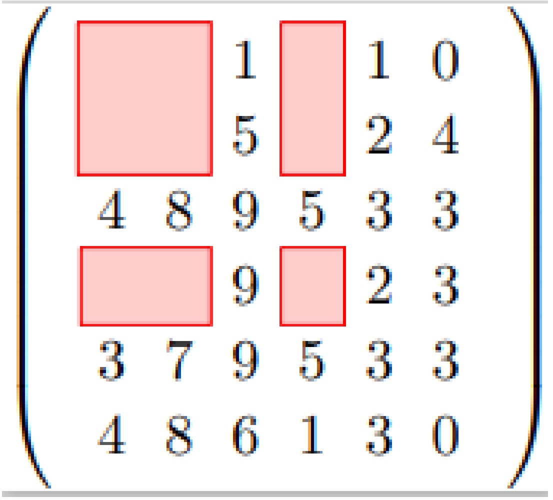


Other applications

Determinantal point processes (DPP):

(PSD) matrix $L \in \mathbb{R}^{n \times n}$: $\mu_L(S) = \det(L_{S,S})$ for $|S| = k$

Applications: random linear algebra, machine learning (recommender system)



[Anari-Liu-Vuong--FOCS'22]

After $\tilde{O}(nk^{\omega-1})$ preprocessing time, can sample from DPP in $\tilde{O}(k^{\omega})$ time.

Other applications

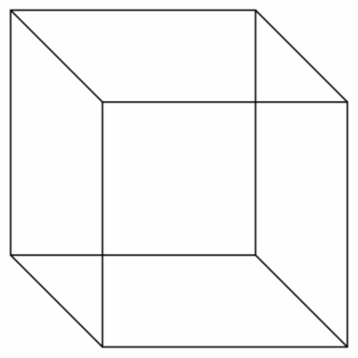
General strongly Rayleigh distributions:

$\mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is strongly Rayleigh if $f_\mu \neq 0$ when $Im(z_i) \geq 0$

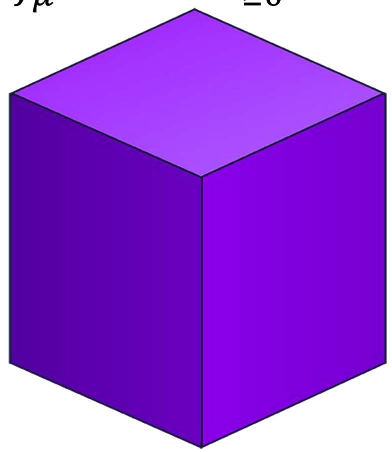
$$f_\mu(z_1, \ldots, z_n) := \sum \mu(S) \prod_{i \in S} z_i$$

Results apply for general strongly Rayleigh distributions, including random spanning tree and DPPs are strongly Rayleigh

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



$f_\mu: \mathbb{R}^n \to \mathbb{R}_{\geq 0}$



Overview

1. Motivation
    - Random spanning trees
    - Determinantal point processes
2. Algorithm
    - Isotropic transformation
    - Up-down walk
3. Analysis

Isotropic transformation

Intuition: make all edges/elements having the same marginal

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

Isotropic transformation

Intuition: make all edges/elements having the same marginal

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

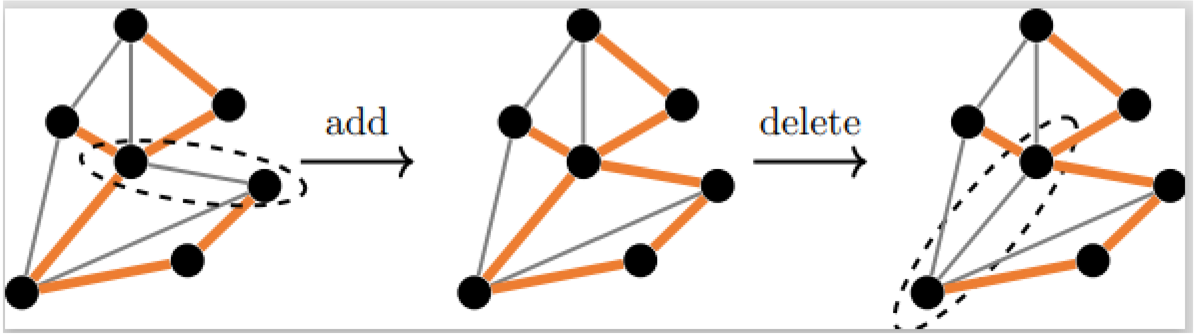
$$\mu: \binom{[n]}{k} \to \mathbb{R}_{\geq 0} \Rightarrow \mu': \binom{U}{k} \to \mathbb{R}_{\geq 0}$$

1. Near isotropy: for $e' \in U$, $Pr_{\mu'}[e' \in T] \leq O(\frac{k}{|U|})$
2. Linear ground set size: $|U| \leq 2n$
3. Preserve strongly Rayleigh property

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 ≡ down-up walk on the complement $\bar{\mu}: \binom{[n]}{n-k} \to \mathbb{R}_{\geq 0}$

defined by $\bar{\mu}([\mathrm{n}] \setminus S) = \mu(S)$

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
- Without isotropy, need $\theta(|E| \log|E|)$ time to converge
- With isotropy, converge in $O(|V| \log|V|)$ time

Overview

1. Motivation
   - Random spanning trees
   - Determinantal point processes
2. Algorithm
   - Isotropic transformation
   - Up-down walk
3. Analysis
   - Improved entropic independence under uniform marginals
   - Mixing time bound using average case local-to-global argument

Entropic independence

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

$\mu\ is\ \frac{1}{\alpha}$-entropic independence ⇔ ∀ν:

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

Entropic independence

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

μ *is* $\frac{1}{\alpha}$-entropic independence ⇔ ∀ν:

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

Strongly Rayleigh ⇒ 1-entropic independence

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

Improved entropic independence under uniform marginals

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

$$\mathcal{D}_{KL}(\bar{\nu}||\bar{\mu}) \geq (n-k)\log(n/k)\,\mathcal{D}_{KL}\left(\nu D_{(n-k)\to 1}\middle|\middle|\mu D_{(n-k)\to 1}\right)$$

Improved entropic independence under uniform marginals

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

$$\mathcal{D}_{KL}(\bar{\nu}||\bar{\mu}) \geq (n-k)\log\left(\frac{n}{k}\right)\mathcal{D}_{KL}(\bar{\nu}D_{(n-k)\to 1}||\bar{\mu}D_{(n-k)\to 1})$$

1. $\mathcal{D}_{KL}(\bar{\nu}||\bar{\mu}) = \mathcal{D}_{KL}(\nu||\mu) \geq k\mathcal{D}_{KL}(\nu D_{k\to 1}||\mu D_{k\to 1})$

Improved entropic independence under uniform marginals

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

$$\mathcal{D}_{KL}(\bar{\nu}||\bar{\mu}) \geq (n-k)\log\left(\frac{n}{k}\right)\mathcal{D}_{KL}\left(\bar{\nu}D_{(n-k)\to 1}\middle|\middle|\bar{\mu}D_{(n-k)\to 1}\right)$$

1. $\mathcal{D}_{KL}(\bar{\nu}||\bar{\mu}) = \mathcal{D}_{KL}(\nu||\mu) \geq k\mathcal{D}_{KL}(\nu D_{k\overset{n}{\to}1}||\mu D_{k\to 1})$

2. $k\mathcal{D}_{KL}(\nu D_{k\to 1}||\mu D_{k\to 1}) \geq (n-k)\log\left(\frac{n}{k}\right)\mathcal{D}_{KL}\left(\bar{\nu}D_{(n-k)\to 1}\middle|\middle|\bar{\mu}D_{(n-k)\to 1}\right)$

Here we use the uniform marginal assumption.

Local to global argument

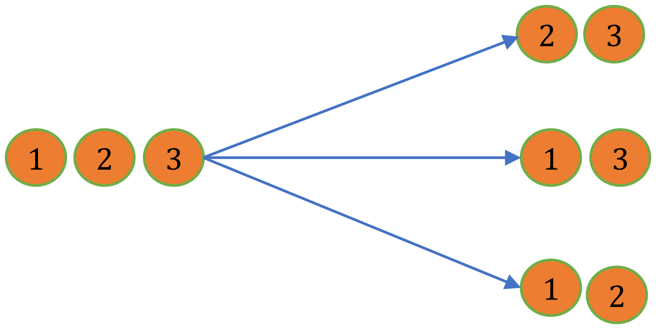
Entropy contraction of $D_{(n-k)\to 1}$ for $\bar{\mu}$ and its conditionals

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

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

Conditional of $\bar{\mu}$ at $\bar{S}$: $\bar{\mu}_{\bar{S}}(\bar{T} \setminus \bar{S}) = \bar{\mu}(\bar{T}|\bar{S}) = \bar{\mu}(\bar{T})$ for $\bar{T} \supseteq \bar{S}$

$D_{(n-k)\to(n-k-1)}$



Local to global argument

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

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

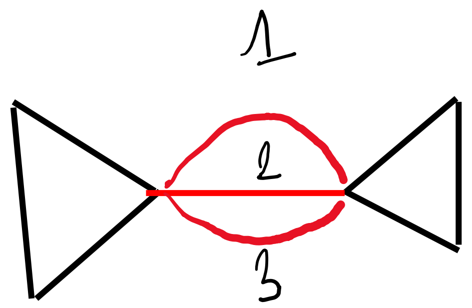
$\Rightarrow$ Mixing time of up-down walk.

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

$$(n-k)\log(\frac{n}{k}) \text{ contraction} \Rightarrow \mathrm{k}\log n \text{ mixing time} ☺$$

But, not all conditionals of $\bar{\mu}$ has improved entropy contraction ☹
i.e. exists $\bar{S}$ s.t.

$$\mathcal{D}_{KL}(\bar{\nu}_{\bar{S}} || \bar{\mu}_{\bar{S}}) < (n-k)\log\binom{n}{k}\mathcal{D}_{KL}(\bar{\nu}_{\bar{S}} D_{(n-k)\to 1} || \bar{\mu}_{\bar{S}} D_{(n-k)\to 1})$$



$\bar{S} = \{\bar{1}, \bar{3}\}$

Average local to global

For each set $\overline{W} \in \binom{[n]}{n-k-1}$ and s, if for "many" $\bar{S} \in \binom{W}{n-s}$

$\bar{\mu}_{\bar{S}}$ has uniform marginal thus improved entropy contraction

then we still get k $\log n$ mixing time ☺

Average local to global

For each set $\overline{W} \in \binom{[n]}{n-k-1}$ and s, if for "many" $\bar{S} \in \binom{W}{n-s}$
$\bar{\mu}_{\bar{S}}$ has uniform marginal thus improved entropy contraction
then we still get $\mathrm{k}\log n$ mixing time ☺
"many" $=$ w/ prob. $1 - 1/n^{10}$ over uniformly chosen $\bar{S}$

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

Proof:

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

Since μ is strongly Rayleigh, marginal doesn't change much

Use martingale argument and Bernstein ineq.

Open problem

- Sublinear sampling alg. for uniform distribution over matroid bases? (log-concave but not necessarily strongly Rayleigh)