Domain Sparsification of Discrete Distributions using Entropic Independence

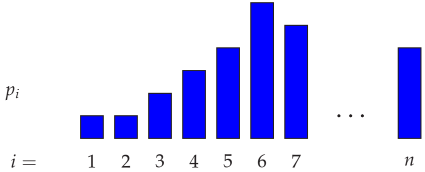
Nima Anari   Michal Dereziński   Thuy-Duong Vuong   Elizabeth Yang

ITCS 2022

November 19, 2022

Given $p_1, \cdots, p_n \geq 0$.
Draw element from $\{1, \cdots, n\}$ with $\mathbb{P}[\text{drawing } i] \propto p_i$.

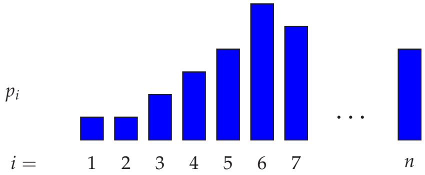
np.random.choice(n, $\underbrace{\text{weight}}_{[p_1,\cdots,p_n]}, \underbrace{\text{cumulative-weight*}}_{[p_1, p_1+p_2, \cdots, p_1+\cdots+p_n]}$, num-sample*)



Given $p_1, \cdots, p_n \geq 0$.
Draw element from $\{1, \cdots, n\}$ with $\mathbb{P}[\text{drawing } i] \propto p_i$.

np.random.choice(n, $\underbrace{\text{weight}}_{[p_1,\cdots,p_n]}$, $\underbrace{\text{cumulative-weight*}}_{[p_1,p_1+p_2,\cdots,p_1+\cdots+p_n]}$, num-sample*)

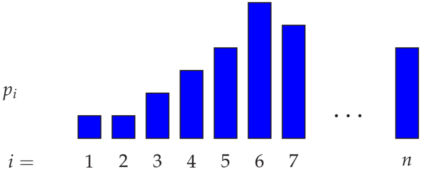
With cumulative-weight, can reduce runtime from $\theta(n)$ to $\theta(\log n)$.



Given $p_1, \cdots, p_n \geq 0$.
Draw element from $\{1, \cdots, n\}$ with $\mathbb{P}[\text{drawing } i] \propto p_i$.

np.random.choice(n, $\underbrace{\text{weight}}_{[p_1,\cdots,p_n]}$, $\underbrace{\text{cumulative-weight*}}_{[p_1, p_1+p_2, \cdots, p_1+\cdots+p_n]}$, num-sample*)

With cumulative-weight, can reduce runtime from $\theta(n)$ to $\theta(\log n)$.
Can we do the same for μ with $\sup(\mu) = n^{\omega(1)}$?



Sampling from discrete distributions

Problem:

Given oracle access to $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, approximately sample:

$$S \sim \mu : \qquad \mathbb{P}[S] \propto \mu(S).$$

ϵ-approximate sample: sample $S \sim \mu'$ s.t. $d_{TV}(\mu, \mu') \leq \epsilon$.

Sampling from discrete distributions

Problem:

Given oracle access to $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, approximately sample:

$$S \sim \mu : \qquad \mathbb{P}[S] \propto \mu(S).$$

ϵ-approximate sample: sample $S \sim \mu'$ s.t. $d_{TV}(\mu, \mu') \leq \epsilon$.

Domain Sparsification

Reducing the task of sampling from μ on $\binom{[n]}{k}$ to sampling from a related distribution ν on $\binom{T}{k}$ such that:

$$|T| \ll n.$$

Counting by repeated sampling

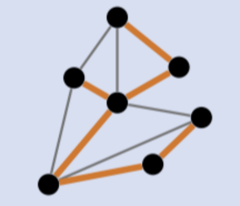
Approximate count $|\Omega|$: Compute $\hat{Z}$ s.t. $\hat{Z} \leq \Omega \leq (1+\epsilon)\hat{Z}$

- μ is uniform over $\Omega \subseteq \binom{[n]}{k}$
- Approximate sampling from $\mu \Leftrightarrow$ Approximate counting $|\Omega|$
  Jerrum-Valiant-Vazirani'86
- To count $|\Omega|$, need to produce many samples from μ. We want to reduce the amortized time-complexity per sample.

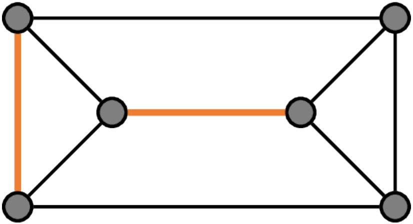
Examples

Determinantal point processes

Spanning trees / forests



k-Matchings



Given: Matrix $L \in \mathbb{R}^{n\times n}$

$$\mathbb{P}[S] \propto \det(L_S) \text{ when } |S| = k$$

Example: $S = \{1,2,4\}$

$$\begin{pmatrix} & & 1 & & 1 & 0 \\ & & 5 & & 2 & 4 \\ 4 & 8 & 9 & 5 & 3 & 3 \\ & & 9 & & 2 & 3 \\ 3 & 7 & 9 & 5 & 3 & 3 \\ 4 & 8 & 6 & 1 & 3 & 0 \end{pmatrix}$$

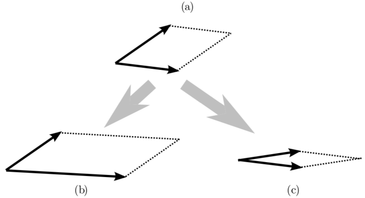
Application:

- Recommender system *Gillenwater-Kulesza-Taskar'12*
- Image Search *Kulesza-Taskar'11*
- RandNLA *Dereziński-Mahonet–AMS Notices'21*

When $L = L^\intercal$, can write $L = VV^\intercal$ where

$$V = \begin{bmatrix} -v_1- \\ \vdots \\ -v_n- \end{bmatrix}, \text{ then}$$

$\mathbb{P}[S] \propto \text{Volume}^2(\text{vectors indexed by } S)$



Gartrell et al'19,20 consider nonsymmetric DPP ($L + L^\intercal \succeq 0$) for its enhanced modelling power.

Outline

1. Background
2. Isotropy for discrete distributions and isotropic transformation
3. Intermediate sampling
4. Domain sparsification: $n \to n^{1-\alpha}\mathsf{poly}(k)$

A brief history

$\underline{n \to \mathsf{poly}(k)}$

Need: negative correlation

- Volume sampling
- (Symmetric) determinantal point processes

Durfee-Peebles-Peng-Rao'17; *Dereziński* [COLT 2019]; *Calandriello- Dereziński-Valko* [NeurIPS 2019, 2020]

A brief history

$\underline{n \to \mathsf{poly}(k)}$

Need: very strong HDX

- Counting matroid bases
- Sampling from log-concave distributions

$\underline{n \to \mathsf{poly}(k)}$

Need: negative correlation

- Volume sampling
- (Symmetric) determinantal point processes

Anari-Dereziński [FOCS 2020]

A brief history

$n \to n^{1-\alpha}\text{poly}(k)$

Minimal assumption: entropic independence

- Counting planar matchings
- Sampling from non-symmetric DPPs

$n \to \text{poly}(k)$

Need: very strong HDX

- Counting matroid bases
- Sampling from log-concave distributions

$n \to \text{poly}(k)$

Need: negative correlation

- Volume sampling
- (Symmetric) determinantal point processes

Anari-Dereziński-V-Yang [arXiv 2021]

Outline

Isotropy for discrete distributions

Definition (*Anari-Dereziński–FOCS'20*)

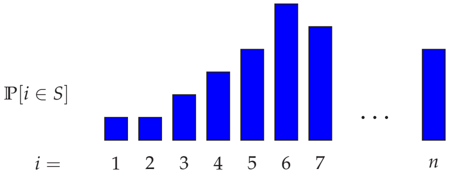
Density $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is (nearly) isotropic if $\mathbb{P}_{S \sim \mu}[i \in S]$ are (nearly) the same for all $i \in [n]$.

Isotropy for discrete distributions

Definition (*Anari-Dereziński–FOCS'20*)

Density $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is (nearly) isotropic if $\mathbb{P}_{S \sim \mu}[i \in S]$ are (nearly) the same for all $i \in [n]$.

Transformation to isotropic position:

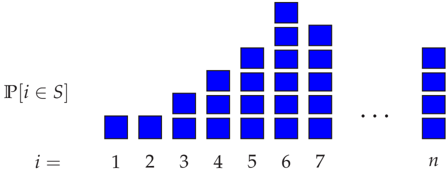


Isotropy for discrete distributions

Definition (*Anari-Dereziński–FOCS'20*)

Density $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is (nearly) isotropic if $\mathbb{P}_{S\sim\mu}[i \in S]$ are (nearly) the same for all $i \in [n]$.

Transformation to isotropic position:



Intuition: Duplicate elements proportionally to the marginals.

Converting to isotropic position

Goal: Efficiently estimate all marginals $\mathbb{P}_{S\sim\mu}[i \in S]$

Converting to isotropic position

Goal: Efficiently estimate all marginals $\mathbb{P}_{S\sim\mu}[i \in S]$

- For DPPs: Marginals (leverage score/effective resistant) are computable in polynomial time

Converting to isotropic position

Goal: Efficiently estimate all marginals $\mathbb{P}_{S\sim\mu}[i \in S]$

- For DPPs: Marginals (leverage score/effective resistant) are computable in polynomial time
- In general: Approximate estimate the marginals by sampling and counting/sampling equivalence *Jerrum-Vazirani-Vazirani'84*
  - Matroid bases
  - Fractionally log-concave distributions (nonsymmetric DPPs, planar matchings)

Converting to isotropic position

Goal: Efficiently estimate all marginals $\mathbb{P}_{S\sim\mu}[i \in S]$

- For DPPs: Marginals (leverage score/effective resistant) are computable in polynomial time
- In general: Approximate estimate the marginals by sampling and counting/sampling equivalence *Jerrum-Vazirani-Vazirani'84*
  - Matroid bases
  - Fractionally log-concave distributions (nonsymmetric DPPs, planar matchings)
- Can speed-up estimating the marginals via an annealing process (see *Anari-Dereziński–FOCS'20*)

How can we use isotropy to accelerate sampling?

Outline

Simple idea: Intermediate sampling

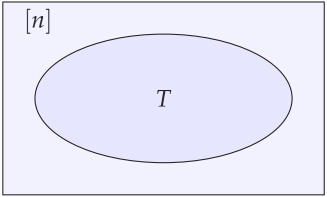
Assumption: μ is nearly isotropic

Simple idea: Intermediate sampling

Assumption: μ is nearly isotropic

1. Intermediate uniform sample:

$$T \sim \binom{[n]}{t}, \qquad k < t < n$$



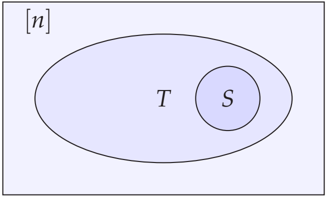
Simple idea: Intermediate sampling

Assumption: μ is nearly isotropic

1. Intermediate uniform sample:

$$T \sim \binom{[n]}{t}, \qquad k < t < n$$

2. Downsample: $S \sim \mu_{|T}$



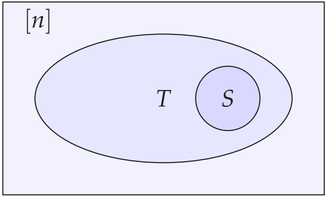
Simple idea: Intermediate sampling

Assumption: μ is nearly isotropic

1. Intermediate uniform sample:

$$T \sim \binom{[n]}{t}, \qquad k < t < n$$

2. Downsample: $S \sim \mu_{|T}$



Domain sparsification from $[n]$ *to* T*?*

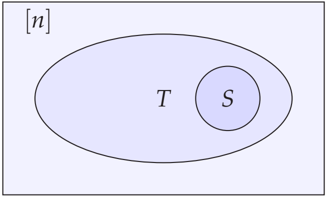
Simple idea: Intermediate sampling

<u>Assumption:</u> μ is nearly isotropic

1. Intermediate uniform sample:

$$T \sim \binom{[n]}{t}, \qquad k < t < n$$

2. Downsample: $S \sim \mu_{|T}$



Domain sparsification from $[n]$ *to* T? Not quite!

Markov chain intermediate sampling

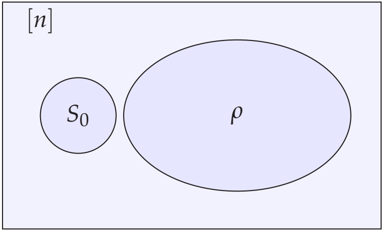
Instantly mixing walk

S_0

Markov chain intermediate sampling

Instantly mixing walk

S_0 $\langle \rho_1, \ldots, \rho_t \rangle$



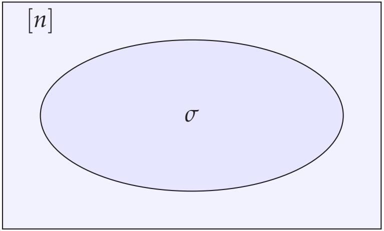
Markov chain intermediate sampling

Instantly mixing walk

$S_0 \qquad \langle \rho_1, \ldots, \rho_t \rangle$

$\searrow \quad \swarrow$

$\langle \sigma_1, \ldots, \sigma_t, \ldots, \sigma_{t+k} \rangle$



Instantly mixing walk

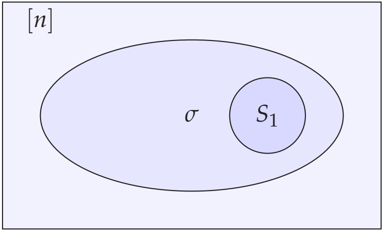
S_0 $\langle \rho_1, \ldots, \rho_t \rangle$

$\searrow$ $\swarrow$

$\langle \sigma_1, \ldots, \sigma_t, \ldots, \sigma_{t+k} \rangle$

$\downarrow$

$S_1 \sim \mu_\sigma$



Markov chain intermediate sampling

Instantly mixing walk

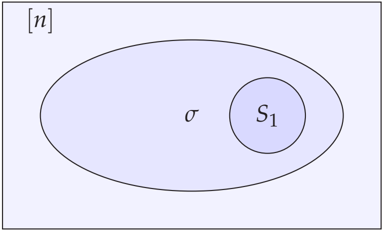
S_0 $\quad \langle \rho_1, \ldots, \rho_t \rangle$

$\searrow \quad \swarrow$

$\langle \sigma_1, \ldots, \sigma_t, \ldots, \sigma_{t+k} \rangle$

$\downarrow$

$S_1 \sim \mu_\sigma$



Hierarchical walk to address generating $S_1 \sim \mu_\sigma$

Markov chain intermediate sampling

Instantly mixing walk

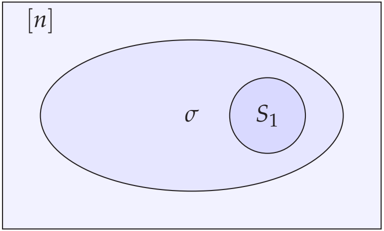
$$S_0 \qquad \langle \rho_1, \ldots, \rho_t \rangle$$

$$\searrow \quad \swarrow$$

$$\langle \sigma_1, \ldots, \sigma_t, \ldots, \sigma_{t+k} \rangle$$

$$\downarrow$$

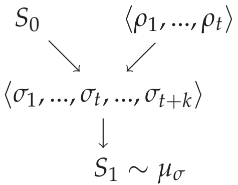
$$S_1 \sim \mu_\sigma$$

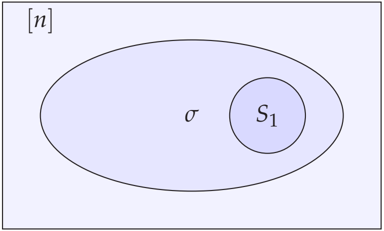


Hierarchical walk to address generating $S_1 \sim \mu_\sigma$

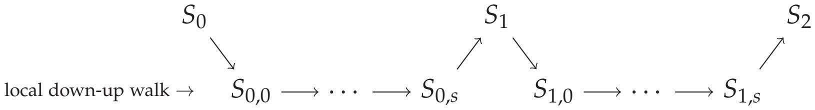
Markov chain intermediate sampling

Instantly mixing walk





Hierarchical walk to address generating $S_1 \sim \mu_\sigma$



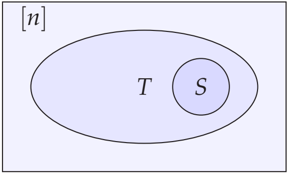
Alimohammadi-Anari-Shiragur-V–STOC'21

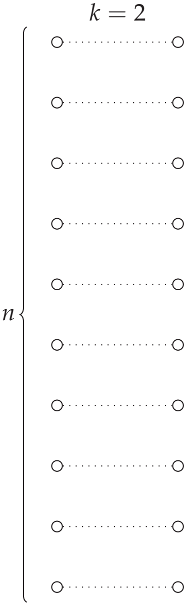
Outline

Example: Bipartite graph

For graph $G = G([n], E)$, consider μ uniform over $E \subseteq \binom{[n]}{2}$

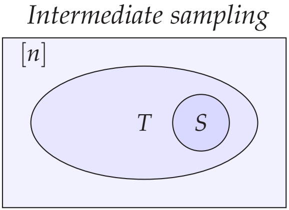
Intermediate sampling





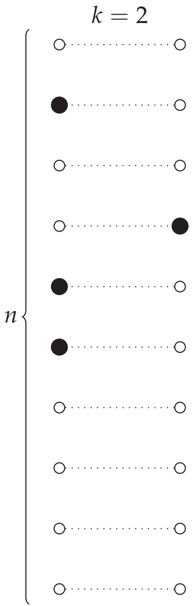
Example: Bipartite graph

For graph $G = G([n], E)$, consider μ uniform over $E \subseteq \binom{[n]}{2}$



$$t := |T| = o(\sqrt{n}):$$

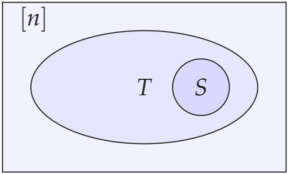
$$\binom{T}{k} \cap \text{supp}(\mu) = \emptyset \text{ almost surely}$$



Example: Bipartite graph

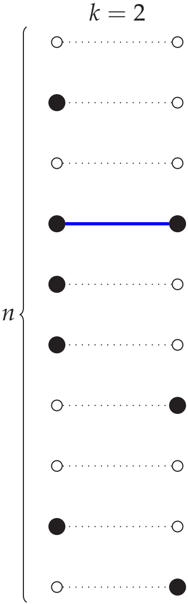
For graph $G = G([n], E)$, consider μ uniform over $E \subseteq \binom{[n]}{2}$

Intermediate sampling



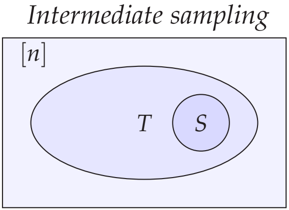
$t = \sqrt{n}$:

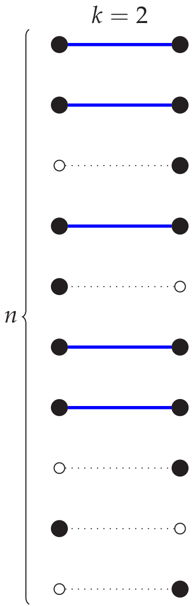
$$\binom{T}{k} \cap \mathrm{supp}(\mu) \neq \emptyset \text{ w. prob. } 1/n$$



Example: Bipartite graph

For graph $G = G([n], E)$, consider μ uniform over $E \subseteq \binom{[n]}{2}$





In general: Birthday paradox for k-collisions

- Generally, for $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, we need $t \simeq n^{1-1/k}$ intermediate samples to find a set $S \in \text{supp}(\mu)$.
  Think: μ encodes a hypergraph

- When can we make $t = n^{0.9}$?

Generating polynomial

For distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, let its **generating polynomial** be

$$f_\mu(z_1, \ldots, z_n) = \sum_S \mu(S) z^S = \sum_S \mu(S) \prod_{i \in S} z_i$$

$f_\mu \equiv$ extension of μ on $\mathbb{R}^n$.

Entropic independence *Anari-Jain-Koehler-Pham-V'21*

Distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is $(1/\alpha)$-*entropically-independent* for $\alpha \in (0,1]$ if, let $p_i = \mathbb{P}_{S\sim\mu}[i \in S]/k$:

Definition 1: Geometry of polynomial

For all $z_i \geq 0$

$$f_\mu(z_1^\alpha, \cdots, z_n^\alpha)^{\frac{1}{\alpha k}} \leq \sum_{i=1}^{n} p_i z_i$$

Definition 2: Entropy contraction

For all distribution $\nu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$

$$\mathcal{D}_{\mathrm{KL}}(\text{marginal of } \nu \mid \text{marginal of } \mu) \leq \frac{1}{\alpha k} \mathcal{D}_{\mathrm{KL}}(\nu \mid \mu).$$

Entropic independence: examples

$1/\alpha$-entropic independence with higher $\alpha \leftrightarrow$ stronger assumption

- Every distribution is $1/\alpha$-entropic independence with $\alpha \geq 1/k$ because by AM-GM

$$f_\mu(z_1^{1/k}, \cdots, z_n^{1/k}) = \sum \mu(S)(\prod_{i \in S} z_i)^{1/k} \leq \sum \mu(S) \frac{\sum_{i \in S} z_i}{k}$$

Entropic independence: examples

$1/\alpha$-entropic independence with higher $\alpha \leftrightarrow$ stronger assumption

- Every distribution is $1/\alpha$-entropic independence with $\alpha \geq 1/k$ because by AM-GM

$$f_\mu(z_1^{1/k}, \cdots, z_n^{1/k}) = \sum \mu(S)(\prod_{i \in S} z_i)^{1/k} \leq \sum \mu(S) \frac{\sum_{i \in S} z_i}{k}$$

- $\alpha = 1$: Spanning tree/symmetric DPPs/uniform distribution over matroid bases
- $\alpha = 1/4$: k-matchings in graph, non-symmetric DPPs

Entropic independence: examples

$1/\alpha$-entropic independence with higher $\alpha \leftrightarrow$ stronger assumption

- Every distribution is $1/\alpha$-entropic independence with $\alpha \geq 1/k$ because by AM-GM

$$f_\mu(z_1^{1/k}, \cdots, z_n^{1/k}) = \sum \mu(S)(\prod_{i \in S} z_i)^{1/k} \leq \sum \mu(S) \frac{\sum_{i \in S} z_i}{k}$$

- $\alpha = 1$: Spanning tree/symmetric DPPs/uniform distribution over matroid bases
- $\alpha = 1/4$: k-matchings in graph, non-symmetric DPPs
- Isotropic transformation preserve entropic independence

Domain sparsification: A general framework

Theorem

Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be $(1/\alpha)$-entropically independent.

Suppose that we have an algorithm $\mathcal{A}$ that can produce approximate samples from any external field λ applied to μ in time $T(m, k)$, where m is the sparsity of λ. Then, we can:

1. *convert μ to nearly-isotropic position in time*
$$O\Big(T(n,k) + n \cdot \mathsf{poly}(k, \log n) \cdot T(n^{1-\alpha}\mathsf{poly}(k), k)\Big).$$
2. *approximately sample from a nearly-isotropic μ in time*
$$O\Big(T(n^{1-\alpha}\mathsf{poly}(k), k)\Big).$$

Application

- Approximate counting size-k matchings in planar graph in time $O(\mathsf{poly}(k)n^2 + \mathsf{poly}(k)n^{3/2}\epsilon^{-2})$
  i.e. output $\hat{Z}$ s.t. $\hat{Z} \le \#\text{matching} \le (1+\epsilon)\hat{Z}$
- After $\tilde{O}(nk^2 + k^3)$ pre-processing, producing sample from
  - Nonsymmetric k-DPP: $\tilde{O}(\mathsf{poly}(k)n^{3/2})$ time.
    For rank-d kernel: $\tilde{O}(\mathsf{poly}(k)n^{3/4}d^2)$-time
  - Symmetric k-DPP: $\tilde{O}(\mathsf{poly}(k))$ time.
    Anari-Liu-V'21 (subsequent work): $\tilde{O}(k^\omega)$ per sample w/ $\tilde{O}(nk^{\omega-1})$ preprocessing.
- For any distribution: after pre-processing, reducing domain size from n to $n^{1-1/k}$.
  This matches birthday-paradox threshold.

Domain size is tight!

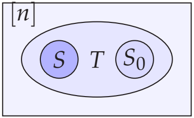
Can "higher-order marginals" $\mathbb{P}_{S\sim\mu}[I \subseteq S]$ help?

Theorem

For any $\alpha \in (0,1]$ and large enough n, k, there is a distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ such that:

1. *μ satisfies $(1/\alpha)$-entropic independence; and*
2. *any domain sparsification scheme to sample from μ requires $t = \tilde{\Omega}(n^{1-\alpha})$, even when given higher-order marginals.*

High level proof idea



Intermediate sampling Markov chain: $S_0 \to S_1 \to \cdots \twoheadrightarrow \mu$

- Prove intermediate sampling Markov chain has instant-mixing: $\mathbb{P}[S_1] \geq \mu(S)(1-0.1)$.
- Lower bound $\frac{\mathbb{P}[S_1]}{\mu(S)}$ by $\underbrace{(\mathbb{E}_{T' \sim \binom{[n]\setminus R}{t-2k}}[\sum_{S' \subset (T' \cup R)} \mu(S')])^{-1}}_{(*)}$,

  with $R = S \cup S_0$.

  Rewrite

$$(*) = (t/n)^k f_\mu(\cdots, \underbrace{(n/t)}_{i \in R}, \cdots, \underbrace{1}_{i \notin R})$$

$$\underset{EI}{\preceq} (t/n)^k \exp(\alpha k \sum_{i \in (S_0 \cup S)} \underbrace{p_i}_{\frac{1}{n}} (n/t)^{1/\alpha}).$$

Conclusions

- Domain sparsification: A general paradigm for reducing the complexity of (repeated) sampling
- Enables high-precision counting for many problems, e.g., counting k-matching in planar graph, size k forests etc.
- Our work generalizes *Anari-Dereziński–FOCS'20*'s domain sparsification framework.

Open problem

Reduce domain size to $\mathsf{poly}(k)$ given higher-order marginals, when μ is α-fractionally log-concave

- α-FLC is strictly stronger assumption than α-entropic independence
- Another way to generalize *Anari-Dereziński–FOCS'20*
- Applications: for nonsymmetric DPPs, partition constraint DPP, k-matchings.

References I

Thank you!