# Privately Counting Partially Ordered Data

Google Research

Matthew Joseph, Mónica Ribero, Alex Yu {mtjoseph, mribero, alexjyu}@google.com

## Background

### Partially Ordered Data

- A poset (P, ≤) is a finite set P together with a reflexive, transitive, and anti-symmetric relation ≤.
- Example: Consider the survey:
  - $x_1$: Have you been told you have hypertension?
  - $x_2$: Have you been told multiple times?
  - $x_3$: Have you had high cholesterol?

  The survey structure imposes partial order $x_2 \leq x_1, x_3$.
- Other examples of partially ordered data include software library dependencies, coursework prerequisites, and any data encoded using directed acyclic graphs.

### Differential Privacy and the K-norm mechanism

**Definition :** Given a dataset $D = \{d_1, \ldots, d_n\} \in \mathcal{D}$ , a randomized mechanism satisfies ε-differential privacy (DP) if for any pair of databases (D, D′) differing in one record, and any set of outputs

$$\mathbb{P}(\mathcal{M}(D) \in S) \leq e^{\epsilon}\mathbb{P}(\mathcal{M}(D') \in S)$$

**Lemma 1 [2]:** Given norm ||·||, statistic $T$ with ||·||-sensitivity Δ , database X, the K-norm mechanism satisfies ε-DP and has output density

$$f_X(y) \propto \exp\left(-\tfrac{\epsilon}{\Delta} \cdot \|y - T(X)\|\right)$$

**Lemma 2 [2]:** Running the K-norm mechanism reduces to sampling the unit ball for the norm ||·||.

### Related Work

- Projection [15, 13], matrix [10, 11], and factorization [4, 14] offer general but potentially slow approximations for optimal, problem-specific private additive noise. We provide a problem-specific, optimal, and fast alternative.
- Joseph and Yu [3] efficiently implemented the K-norm mechanism for improved noise distributions in sum, count, and vote problems with contribution bounds.
- Our work differs from Joseph and Yu [7] by addressing different problems needing different sampling techniques due to their unique underlying combinatorial structures and the resulting challenges in sampling their respective norm balls.

## Contributions and Methods

**For partially ordered data, tailoring an instance of the K-norm mechanism produces a more accurate algorithm that's still fast enough to be practical.**

### Step 1: relate the poset ball to a known combinatorial object.

**Lemma 3.4.** *The poset ball for poset $(P^*, \preceq)$ is $\mathcal{O}_2(P^* - r)$, the double order polytope on the double order poset $(P^* - r, \preceq, \preceq)$.*

### Step 2: Efficiently sample the "known" object:

Idea:

1. Find ***unimodular triangulation*** of the polytope.
   - Disjoint subdivision of the polytope into minimal volume simplices.
   - *Extended bipartitions* are partitions of the nodes in two sets where each part can be extended to a total order that is compatible with the underlying partial order.
   - Extended bipartitions index a triangulation of the polytope.
2. Sample one simplex in the triangulation.
   - Extended bipartitions can be efficiently sampled.
   - The indexing structure can be efficiently translated into vertices of the corresponding simplex.
3. Sample uniformly at random from the corresponding simplex.

**Theorem 3.15.** *The poset ball for $(P^*, \preceq)$ can be sampled in time $O(d^2)$.*

### Why not rejection sampling?

**Theorem 3.17.** *Rejection sampling the poset ball using any $\ell_p$ ball is inefficient.*

## Results

### Error introduced by different norms

Lemma 3.16. Proves the is the $B_\infty^d$ is the minimum $\ell_p$ ball containing the poset ball. Below we empirically evaluate the expected $\ell_2$-norm of a vector sampled from different $\ell_p$ balls.

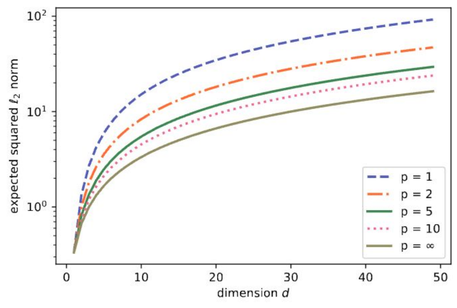
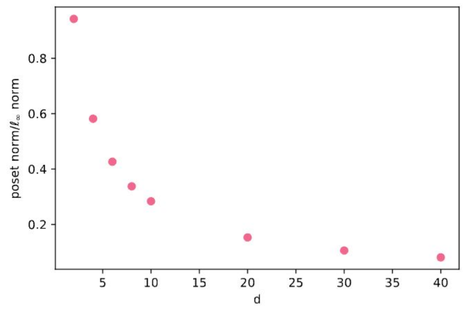



Figure 1: $r_{p,d}^2 \mathbb{E}_2^2(B_p^d)$ (see Lemma 4.1).

### Poset mechanism vs $\ell_\infty$ mechanism

We uniformly sample a directed acyclic graph on a fixed number d of uniquely labeled vertices. our algorithm's advantage in terms of expected squared ℓ2 norm widens with d.




### Error reduction on a real survey

| # survey sections | poset ball squared $\ell_2$ norm / $l_\infty$ ball squared $\ell_2$ norm |
|---|---|
| 1 | 0.414 |
| 2 | 0.427 |
| 3 | 0.408 |

Figure 4: NHIS average mean squared $\ell_2$ norm ratios, from 10,000 trials each.

References

- [1] Center for Medicare Services and Medicaid. National health interview survey. https://www.cms.gov/
- [2] Hardt, Moritz, and Kunal Talwar. "On the geometry of differential privacy. STOC 2010.