Graphical-model based estimation and inference for differential privacy

Ryan McKenna¹, Daniel Sheldon¹,², Gerome Miklau¹
¹University of Massachusetts, Amherst
²Mount Holyoke College
Inference in Privacy Mechanisms

Sensitive Data → Randomized Algorithm → Workload Answers
Inference in Privacy Mechanisms

Sensitive Data $\rightarrow$ Randomized Algorithm $\rightarrow$ Private Observations $\rightarrow$ Inference Algorithm $\rightarrow$ Workload Answers
Inference in Privacy Mechanisms

- Existing techniques for inference either don’t scale or don’t extract the most utility from the private observations.

- Proper inference has many benefits:
  - Resolves inconsistencies
  - Improves utility
  - Answers new queries
  - Supports synthetic data generation
Problem Statement

- Given:
  - an unknown discrete data distribution \( p \in \mathbb{R}^n \)
  - a query matrix \( Q \in \mathbb{R}^{m \times n} \)
- Our observation model is:
  \[
y = Qp + \varepsilon
\]
- We want to recover an estimate of \( p \) from \( y \)
  \[
  \hat{p} \in \arg \min_{p \in S} ||Qp - y||
  \]
- Size of \( p \) is intractably large

Random Laplace or Gaussian noise
Approach

- Reformulate problem to find a graphical model $p_\theta$ instead

$$\hat{\theta} \in \arg\min_\theta ||Qp_\theta - y||$$

Much smaller than $p$

- If $Q$ only depends on $p$ though its marginals,
  - We can solve this problem efficiently
  - Solution to reformulated problem is the maximum entropy solution to the original problem
Scalability Improvements of PGM

- Graphical-model inference scales much better than traditional approaches.

![Graph showing the scalability of PGM compared to traditional approaches.](chart.png)

- Traditional approaches fail at 10 dimensions.
- PGM scales to 1000 dimensions.
Utility Improvements of PGM

- Graphical-model inference improves the utility of several state-of-the-art privacy mechanisms.

We offer similar improvements for DualQuery, HDMM, and MWEM as well (see poster).
Graphical-model based estimation and inference for differential privacy

Poster #171

Code available on GitHub:
https://github.com/ryan112358/private-pgm