Privately Learning Markov Random Fields

Huanyu Zhang, Cornell University
Gautam Kamath, University of Waterloo
Janardhan Kulkarni, Microsoft Research
Zhiwei Steven Wu, University of Minnesota

Table of contents

Problem formulation

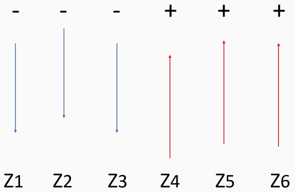
Ising models

$\mathcal{D}(A)$ is a distribution on $\{\pm 1\}^p$ s.t.

$$\Pr(Z = z) \propto \exp\left(\Sigma_{i<j} A_{i,j} z_i z_j + \Sigma_i A_{i,i} z_i\right),$$

where $A \in \mathbb{R}^{p \times p}$ is a **symmetric** weight matrix.

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$



Applications of Ising models

Ising models are heavily used in physics, social network, etc.

Magnet:

- Each dimension represents a particular 'spin' in the material.
- -1 if the spin points down or $+1$ if the spin points up.

Social network:

- Each of the dimensions is a person in the network.
- -1 represents voting for Hilary; $+1$ represents for Trump.

Two alternative objectives

h: unknown Ising model

Input: i.i.d. samples X_1^n from h

Structure learning: output $\hat{A} \in \{0,1\}^{p \times p}$ s.t.

$$\text{w.h.p.,} \quad \forall i \neq j, \hat{A}_{i,j} = \mathbf{1}(A_{i,j} \neq 0).$$

Parameter learning: given accuracy α, output $\hat{A} \in \mathbb{R}^{p \times p}$ s.t.

$$\text{w.h.p.,} \quad \forall i \neq j, \ \left| \hat{A}_{i,j} - A_{i,j} \right| \leq \alpha.$$

Sample complexity: least n to estimate h

Privacy

Data may contain **sensitive** information.

Medical studies:

- Learn behavior of genetic mutations.
- Data contains health records or disease history.

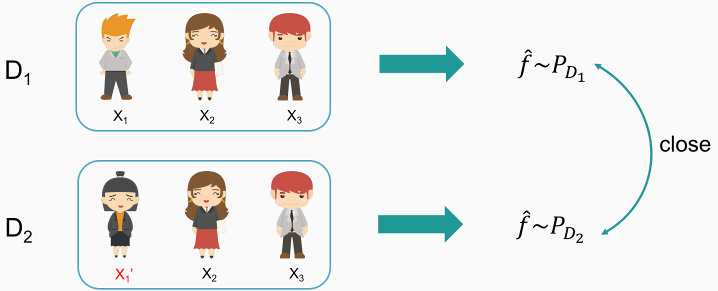
Navigation:

- Suggests routes based on aggregate positions of individuals.
- Position information indicates users' residence.

Differential privacy (DP) [Dwork et al., 2006]

$\hat{f}$ is (ε, δ)-DP for any X_1^n and Y_1^n, with $d_{ham}(X_1^n, Y_1^n) \leq 1$, for all measurable S,

$$\Pr\left(\hat{f}(X_1^n) \in S\right) \leq e^{\varepsilon} \cdot \Pr\left(\hat{f}(Y_1^n) \in S\right) + \delta$$



Privately learning Ising models

Given i.i.d. samples from distribution p, the goals are:

- *Accuracy*: achieve structure learning or parameter learning.
- *Privacy*: estimator must satisfy (ε, δ)-DP.

Main results

Main results

Assumption: the underlying graph has a bounded degree.

	Parameter Learning	**Structure Learning**
Non-private	$O(\log p)$ [Wu et al., 2019]	$O(\log p)$ [Wu et al., 2019]
(ε, δ)-**DP**	$\Theta(\sqrt{p})$	$\Theta(\log p)$
$(\varepsilon, 0)$-**DP**	$\Omega(p)$	$\Omega(p)$

Main results

Assumption: the underlying graph has a bounded degree.

	Parameter Learning	**Structure Learning**
Non-private	$O(\log p)$ [Wu et al., 2019]	$O(\log p)$ [Wu et al., 2019]
(ε, δ)-**DP**	$\Theta(\sqrt{p})$	$\Theta(\log p)$
$(\varepsilon, 0)$-**DP**	$\Omega(p)$	$\Omega(p)$

Only (ε, δ)-DP structure learning is **tractable** in high dimensions!

Private structure learning

Private structure learning - upper bound

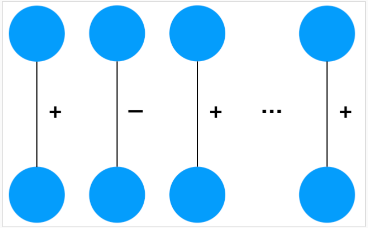
Our (ε, δ)-DP UB comes from **Propose-Test-Release**.

Lemma 1 [Dwork and Lei, 2009]. Given the existence of a m-sample non-private SL algorithm, there exists an (ε, δ)-DP algorithm with the sample complexity $n = O\left(\frac{m \log(1/\delta)}{\varepsilon}\right)$.

We note that this method does not work when $\delta = 0$.

Private structure learning - lower bound

Our $(\varepsilon, 0)$-LB comes from a reduction from **product distribution learning**.



By **packing** argument, we show $n = \Omega(p)$.

Private structure learning

	Parameter Learning	Structure Learning
Non-private	$O(\log p)$ [Wu et al., 2019]	$O(\log p)$ [Wu et al., 2019]
(ε, δ)-**DP**		
$(\varepsilon, 0)$-**DP**		

Private structure learning

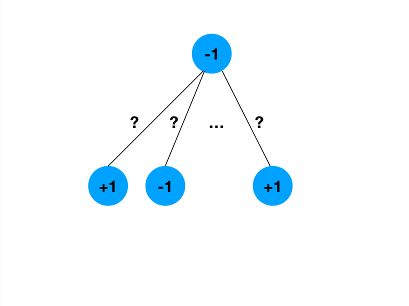
	Parameter Learning	Structure Learning
Non-private	$O(\log p)$ [Wu et al., 2019]	$O(\log p)$ [Wu et al., 2019]
(ε, δ)-**DP**		$\Theta(\log p)$
$(\varepsilon, 0)$-**DP**	$\Omega(p)$	$\Omega(p)$

Private parameter learning

Private parameter learning - upper bound

The following lemma is a nice property of Ising model.

Lemma 2. Let $Z \sim \mathcal{D}(A)$, then $\forall i \in [p]$, $\forall x \in \{\pm 1\}^{[p-1]}$, $\Pr\left(Z_i = 1 | Z_{-i} = x\right) = \sigma(\Sigma_{j \neq i} 2A_{i,j}x_j + 2A_{i,i})$.



Question: Can we utilize **sparse logistic regression**?

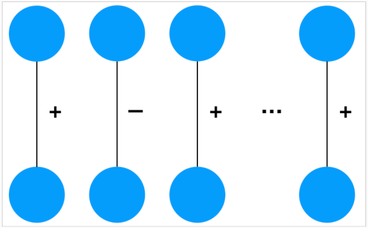
Private parameter learning - upper bound

Answer: Yes! And there are two advantages:

- $O(\log p)$ samples are enough without privacy [Wu et al., 2019].
- It can be **efficiently** and **privately** solved by private Frank-Wolfe algorithm [Talwar et al., 2015].

Private parameter learning - lower bound

We consider a similar reduction as structure learning.



Our (ε, δ)-DP LB comes from a reduction from **product distribution learning**.

Private parameter learning

	Parameter Learning	Structure Learning
Non-private	$O(\log p)$ [Wu et al., 2019]	$O(\log p)$ [Wu et al., 2019]
(ε, δ)-**DP**	$\Theta\left(\sqrt{p}\right)$	$\Theta(\log p)$
$(\varepsilon, 0)$-**DP**	$\Omega(p)$	$\Omega(p)$

Generalization to other GMs

Generalization to other GMs

Similar results are shown in other graphical models:

- Binary t-wise Markov Random Field:
  From pairwise to t-wise dependency.

- Pairwise Graphical Model on General Alphabet:
  Alphabet from $\{\pm 1\}^p$ to $[k]^p$.

The End

Paper ID: 112 Details in paper online:

`https://arxiv.org/pdf/2002.09463.pdf`

Dwork, C. and Lei, J. (2009).
Differential privacy and robust statistics.
In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006).
Calibrating noise to sensitivity in private data analysis.
In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284, Berlin, Heidelberg. Springer.

Talwar, K., Thakurta, A. G., and Zhang, L. (2015).
Nearly optimal private lasso.
In *Advances in Neural Information Processing Systems*, pages 3025–3033.

Wu, S., Sanghavi, S., and Dimakis, A. G. (2019).

Sparse logistic regression learns all discrete pairwise graphical models.
In *Advances in Neural Information Processing Systems*, pages 8069–8079.