The exponential mechanism via MCMC

Aleksandar Makelov, mentored by prof. Salil Vadhan

Harvard College ’15, supported by HCRP

Privacy Tools for Sharing Research Data

A National Science Foundation Secure and Trustworthy Cyberspace Project

Motivation

• The exponential mechanism is a generic differentially-private algorithm, but the **sample space often has exponential size**.

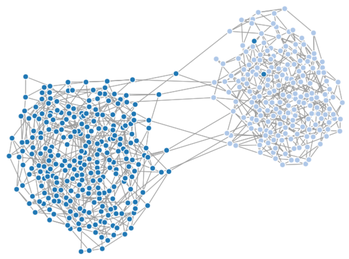
• **Markov Chain Monte Carlo** is a generic technique for **sampling from exponentially large sample spaces**.

• Therefore, maybe there are situations in which MCMC can successfully be applied to differential privacy.

Proof of concept

• The final project for Salil’s spring 2013 course on differential privacy that Arpon Raksit and I completed offered experimental evidence in favor of that approach.

• The problem we worked on was efficiently generating synthetic data that preserves the community structure of a social network and respects edge-level differential privacy.



A synthetic dataset which is 0.3-differentially private

Markov chains and MCMC

•A Markov chain is an abstraction for a discrete memoryless stochastic process.

• It is a central object in probability theory, where it finds diverse applications.

• Moreover, Markov chains are the main objects of study in Markov chain Monte Carlo (MCMC) methods for sampling from probability distributions in which we can efficiently access ratios of pointwise probability masses.

• The distribution associated with the exponential mechanism has this property whenever the utility function is efficiently computable.

• The main idea behind MCMC is to simulate a Markov chain that has the desired distribution as its stationary distribution, and to wait until it converges.

• For an implementation of MCMC to be **efficient,** we need the corresponding Markov chain to be **rapidly mixing,** in other words, converging quickly to its stationary distribution**.**

My project

• I spent the first half of the duration of my research project on studying Markov chains and the general theoretical frameworks for proving rapid mixing.

• As suggested by Salil, I looked into problems from statistical physics, where many of the distributions that arise naturally are of the same form as those coming from the exponential mechanism.

• In particular, I found that the Ising model, a well-known distribution, is closely related to the problem of privately releasing the min-cut of a graph that has been studied in [1]

• In [1], the authors exhibit an (ε, δ)-differentially private mechanism for releasing min-cut, where δ is inversely polynomial in the size of the problem instance.

• One of my main goals was to remove the δ dependence from the privacy. There are two conceptually different sources of δ in the algorithm in [1], and I managed to remove the first.

• In [2], the authors outline a polynomial-time sampling algorithm for the Ising model. This comes very close to removing the second source of δ-privacy in [1], but the distribution

My project, continued

corresponding to the Ising model is not exactly the same as the distribution in the min-cut problem.

• I investigated other well-studied distributions from statistical physics, such as the hardcore model, but they turned out to be too restrictive for the exponential mechanism to achieve good utility.

Conclusions

• The correspondence between the exponential mechanism and well-studied problems in statistical physics seems to be a promising avenue for future research.

• In particular, I conjecture that a better understanding of the ideas behind [2] can lead to an implementation of ε-differentially private min-cut.

• Thanks to Salil and the Harvard College Research Program for making this project happen!

References

[1] Gupta, Anupam, et al. "Differentially private combinatorial optimization." Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2010.

[2] Randall, Dana, and David Wilson. "Sampling spin configurations of an Ising system." Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, 1999.