Inference from the Data: Protect Privacy when Releasing Results

Random noises in algorithms help protect individuals' information from privacy attacks.

Bingyue Su

Abstract

Many companies collect and analyze users' data to learn about and improve the performance of their products. However, users' data often contain users' personal information. When releasing results from an analysis, adversaries may infer an individual's information via privacy attack or threat models. In this project, we develop the privacy-preserving version of Metropolis-Hastings (MH), a widely used sampling algorithm in statistics, and find that the privatized algorithm maintains good data utility in a financial data set and in simulated data sets with formal privacy guarantees.

Why Should Industry Care?

Statistical analysis and machine learning of user and customer data can help companies and financial institutions to evaluate users' satisfaction about a product or a service and provide valuable information and insights on how to improve the quality of the product or service. For example, Netflix may improve its recommendation algorithm by analyzing users' comments; banks can analyze customer data to determine which customers may need a loan or credit card.

Anonymization and pseudonymization such as removing personal identifiers from a released data set is not an effective approach to avoid privacy risk. For example, Netflix launched a recommendation algorithm competition in 2006, publishing the users' data after removing their name ID. However, attackers could still tell an individual's identity by using public Internet Movie Database (IMDB) data. In addition, aggregate results from statistical analysis and machine learning can expose users' personal information upon release.

In summary, randomized algorithms with formal privacy guarantees would be useful in some data or results scenarios to provide effective privacy protection while maintaining utility in the released results perturbed by the randomized algorithms.

Introduction

In some cases, adversaries may be able to infer individuals' sensitive information from released aggregated statistical results. For example, suppose we have a database containing the incomes of employees in a company. If there is only one female employee, people may propose two queries "the average income of male employees" and "the average income of all employees." People could infer the income of the female employee by comparing the results of queries. We refer to this type of privacy attacks "differential attacks." To protect privacy attacks like this, we can use the *differential privacy (DP)* concept (Dwork, 2006).

DP is a state-of-the-art mathematical privacy concept in privacy research and applications. DP provides quantifiable formal privacy guarantees by perturbing results via randomized algorithms before release. Many technology companies have applied differential privacy to their products. For example, Apple uses DP to perturb data shared by users so that they can know the behavior of a user group but cannot infer one individual's information. In the above income example, we would perturb the mean and then release the perturbed version to the public. We could then illustrate the difficulty of inferring private information from the perturbed version under the DP framework.

The key issue in developing and applying DP methods in general is the tradeoff between privacy protection and data utility of released results via randomized algorithms with DP guarantees. This is illustrated by Google's community mobility report during the COVID-19 pandemic, which utilizes DP to transform personal travel records. If privacy protection is too weak, attackers may infer personal information. Conversely, much perturbation may influence the overall pattern and weaken the report's utility.

In general, the smaller the privacy loss, the less the utility there is for the released results. Given a pre-set level of privacy guarantees (referred to as the privacy budget or privacy loss parameter), the aim is to achieve the greatest data utility possible.

In this research project, we improve on the MH algorithm, a widely-used method for generating samples from a probability distribution, by proposing a new privacy-preserving algorithm with DP guarantees while aiming to achieve better utility in the generated samples in comparison to other existing methods for generating privacy-preserving samples from a distribution.

Differentially Private Metropolis-Hastings Sampling through Auxiliary Variables (DP-MHAV)

During the running process of the MH algorithm, a statistician must use the data set thousands of times, which increases the risk of privacy leaks. Inspired by the possible risk of this popular algorithm, we decided to establish a privacy-preserving framework.

There is built-in randomness in the MH algorithm. Wang et. al. (2015) showed that drawing one sample from data distribution provides DP guarantees without extra randomness. Heikkila et. al. 2019) found that the subsampling Metropolis-Hastings algorithm is naturally differentially private.

Our method, Differentially Privacy Metropolis-Hastings Sampling through Auxiliary Variables (DP-MHAV), allows users of the method to adjust their desired level of the privacy budget. In addition, the way we added DP random noise to the MH algorithm differs from previous DP mechanisms for MH algorithms. The form of random noise has been designed to be more compatible with MH algorithms than previous DP approaches.

We also leveraged privacy amplification, a theory which demonstrates that subsampling improves privacy protection (Balle et. al., 2018). In addition, subsampling also reduces the computational burden when running MH algorithms on large datasets, which is typical of user and customer data collected by companies nowadays.

When evaluating the utility of the samples generated by DP-MHAV, we monitored the accuracy of parameter estimation and their statistical inferential properties. To compare the utility of DP-MHAV and other privacy-preserving MH algorithms at the same privacy budgets, we ran experiments on both simulated data and a real financial data set. The simulation experiment used data sets generated from a mixed Gaussian distribution. We examined the estimation and inferred the parameters of the Gaussian distribution. The financial data set contained information on more than 40,000 clients. The goal was to predict whether a client subscribed to a term deposit based on their attributes, such as education background and job type, using a Bayesian logistic model. In addition, we also obtained statistical inferences on the model parameters.

We also encountered issues during the development and implementation of the DP-MHAV algorithm. The main limit of DP-MHAV was its convergence to the target distribution as it is an iterative procedure starting from a random distribution. In addition, with the subsampling and the random perturbation used in each iteration, the algorithm may take more time to generate

reliable samples. Users of the DP-MHAV algorithm must be attentive when adjusting the hyper-parameters used in the algorithm.

Findings

The results from the experiments suggest our approach produced non-inferior statistical inference on model parameters in the simulated data and better statistical inference and more accurate predictions when predicting whether a client has subscribed to a term deposit.

Table 1 presents the results on two parameters θ_1 and θ_2 in the Gaussian mixture model in the simulation study. Privacy budget is the variable measuring the privacy protection, lower values meaning stronger protection. Gaussian and Laplace are basic DP approaches that add Gaussian or Laplace noise to the algorithm directly. WFS'15 is the DP method proposed by Wang, 2015. Since the true values of θ_1 and θ_2 are 0 and 5, we expected the results of DP-MHAV are closest to them. However, it showed that all the methods produce similar estimations, which means they were almost equally favorable in this experiment setting.

Table 1:Simulation Result

	non-private	Privacy-preserving				
Parameter	Original	Privacy budget	DP-MHAV	WFS'15	Gaussian	Laplace
$\theta_1 = 0$	0.0011	0.5	0.0031	0.0011	0.0018	0.0044
		1	0.0025	0.0014	0.0035	0.0009
		2	0.0024	0.0006	0.0015	0.0030
$\theta_2 = 5$	4.9774	0.5	4.9748	4.9773	4.9772	4.9717
		1	4.9749	4.9767	4.9731	4.9757
		2	4.9747	4.9791	4.9771	49759

The experiment results for the financial data set application are shown in Figure 1. Each dot in Figure 1 represents the effect of a predictor on whether a client is subscribed to a term deposit with an error bar (95% posterior interval), and each column shows the results from one method. "Original" refers to the MH algorithm without privacy protection. "Original subset" represents the MH algorithm utilizing the subsampled data set. The other methods are the same as in Table 1. We observed that the privacy-preserving estimates of the effects of the

predictors in the DP-MHAV algorithm were closer to the original estimation compared to other privacy-preserving methods, implying DP-MHAV had the highest utility in this analysis.

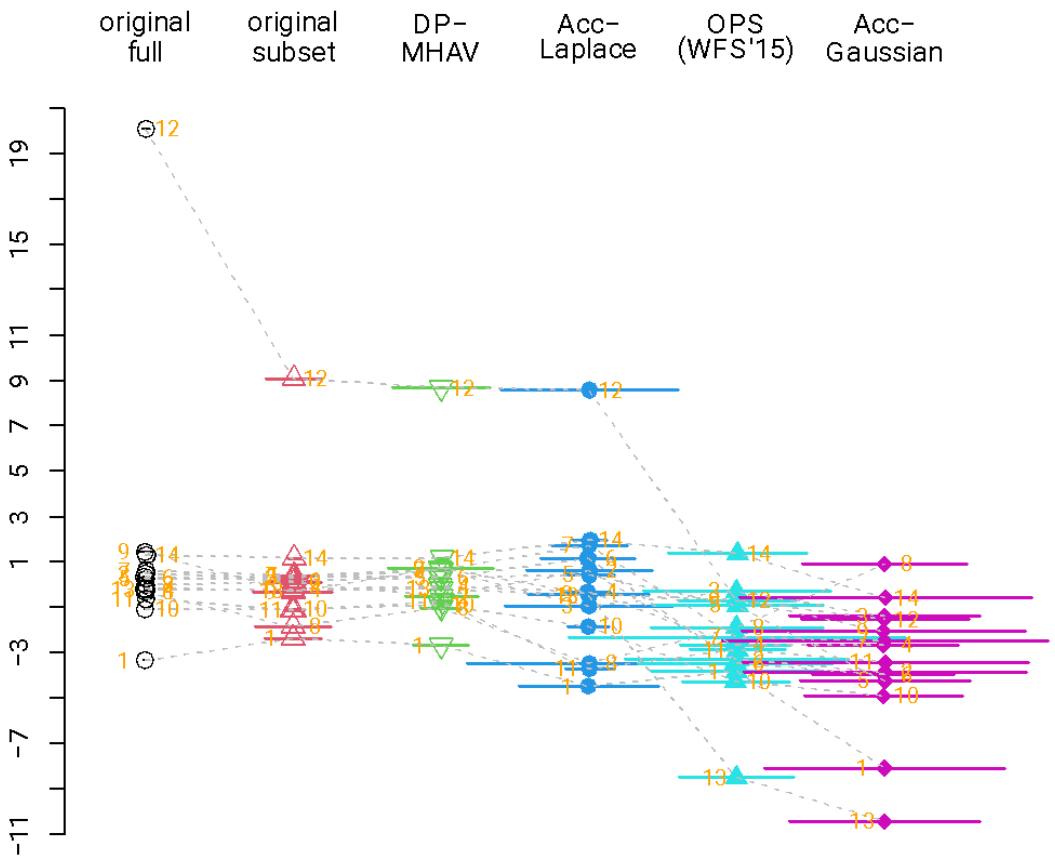

Figure 1:Bank Marketing Data Result

Figure 2 shows the prediction accuracy results on whether a client is subscribed to a term deposit using the privacy-preserving samples of the methods. The closer a line to the upper left corner, the more accurate this method is. The square of the area under a line is the AUC value. Table 2 lists the AUC values for those methods, and DP-MHAV attained the highest AUC among all privacy-preserving methods, implying DP-MHAV provided the most accurate prediction.

Table 2: AUC values

Original Full	Original Subset	DP-MHAV	WFS'15	Gaussian	Laplace
0.869	0.824	0.830	0.617	0.590	0.754

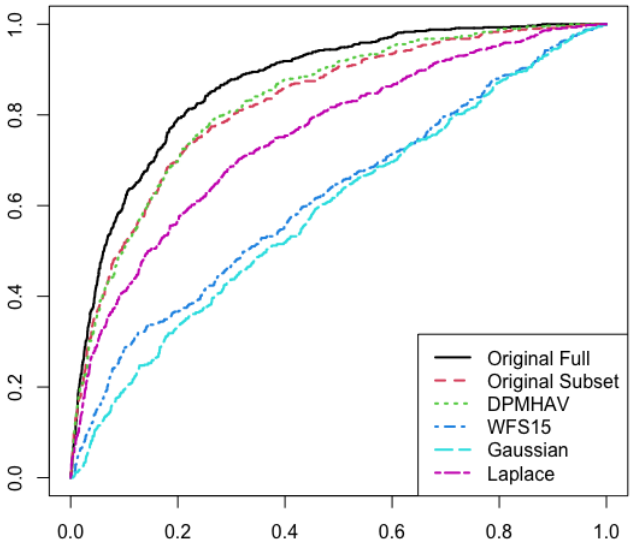

Figure 2: ROC plot

Practical Applications

Our method can benefit user and customer-oriented companies, such as tech companies, financial institutes, and the healthcare industry, which often rely on user and customer data and feedback to improve their service and products while being constantly pressured to protect their privacy when collecting, analyzing, and sharing their data.

For example, the risk department at a bank may use user and customer data to train models to predict whether to issue credit cards and loans to an applicant. Releasing the trained model in this case may incur privacy leaks. The DP-MHAV algorithm can thus be used to provide privacy guarantees for models that involve sampling from a distribution, such as Bayesian logistic regression.

As another example, electronic health records (EHRs) are great data sources for medical informatics that also contain sensitive medical information about individual patients. Due to privacy concerns on medical information, sharing of and access to large single-center EHR data is often infeasible in practice. In this case, our algorithm can be potentially used to generate pseudo EHR datasets to release with formal privacy guarantees.

Conclusion

We developed DP-MHAV, a new DP mechanism, in this research project and proved the corresponding privacy guarantees when releasing samples from a distribution via DP-MHAV.

The preliminary results from the experiment demonstrated the advantage of DP-MHAV compared to previous DP mechanisms in terms of data utility.

Our research also demonstrated that different DP mechanisms can be very different in their utility as well as computational cost in practical implementation. Users of DP methods need to choose proper mechanisms, especially those involving iterations and accessing the original data repeatedly.

For future work, we will continue to solve the issues in the current theoretical analysis of DP-MHAV in terms of its convergence guarantee to a stationary distribution. In addition, the simulation results need improvement. We will refine the experiment setting and generate more empirical evidence regarding the utility of DP-MHAV in the near future.

Bibliography

Balle, B. a.-X. (2018). Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *International Conference on Machine Learning* (pp. 394-403). JLMR.

Dwork, C. a. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Springer.

Heikkila, M. a. (2019). Differentially private markov chain monte carlo. In *Advances in Neural Information Processing Systems* (pp. 4113-4123).

Wang, Y.-X. a. (2015). Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *International Conference on Machine Learning* (pp. 2493-2502).