Template for ACM CCS 2017

Anonymous Author(s)

ABSTRACT

Your abstract should go here. You will also need to upload a plain-text abstract into the web submission form.

CCS CONCEPTS

• **Security and privacy** → Use https://dl.acm.org/ccs.cfm to generate actual concepts section for your paper;

KEYWORDS

template; formatting; pickling

1 PROBLEM STATEMENT

Our research is based on an observation that under different measurement, mechanisms usually have different accuracy behavior. When adding noise according to a certain measurement, mechanism' accuracy behavior will be better than when adding noise according to other measurements. Here we study the private Bayesian inference based on Hellinger distance measurement. Showing that our newly-designed mechanism adding noise according to Hellinger distance performs better than baseline mechanism adding noise according to l_1 norm.

2 TECHNICAL PROBLEM STATEMENT

Our study based on the Bernoulli-Beta Bayesian inference. In the Bayesian inference, first there is a prior distribution $\pi(\boldsymbol{\xi})$ to present our belief about parameter $\boldsymbol{\xi}$. Then, we get some observed data x sizing n, and produce a posterior distribution $Pr(\boldsymbol{\xi}|x)$. The Bayesian inference is based on the Bayes' rule to calculate the posterior distribution:

$$Pr(\boldsymbol{\xi}|x) = \frac{Pr(x|\boldsymbol{\xi})\pi(\boldsymbol{\xi})}{Pr(x)}$$

It is denoted as $\mathsf{BI}(x, \pi(\boldsymbol{\xi}))$ taking an observed data set $x \in \mathcal{X}^n$ and a prior distribution $\pi(\boldsymbol{\xi})$ as input, outputting a posterior distribution of the parameter $\boldsymbol{\alpha}$. For conciseness, when prior is given, we use $\mathsf{BI}(x)$. n is the size of the observed data size.

In our inference algorithm, we take a Dirichlet distribution as prior belief for the parameters, $\mathsf{DL}(\boldsymbol{\alpha})$, where $\pi(\boldsymbol{\xi}) = \mathsf{DL}(\boldsymbol{\xi}|\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{m} \xi_i^{\alpha_i}}{B(\boldsymbol{\alpha})}$, and the Bernoulli distribution as the statistic model for $Pr(x|\boldsymbol{\alpha})$. m is the order of the Dirichlet distribution.

We give a inference process based on a concrete example, where we throw an irregular m sides dice. We want to infer the probability of getting each side $\boldsymbol{\xi}$. We get a observed data set $\{s_{k_1}, s_{k_2}, \cdots, s_{k_n}\}$ by throwing the dice n times, where $k_i \in \{1, 2, \cdots, m\}$ denotes the side we get when we throw the dice the i^{th} time. The posterior distribution is still a Dirichlet distribution with parameters $(\alpha_1 + n_1, \alpha_2 + n_2, \cdots, \alpha_m + n_m)$, where n_i is the appearance time of the side i in total.

In the case that $m = 2$, it is reduced to a Beta distribution $\text{beta}(\alpha, \beta)$. The m side dice change into a irregular coin with side A and side B. The posterior is computed as $(\alpha + n_1, \beta + n_0)$, where n_1 is the appearance time of side A in the observed data set and n_0 is the appearance time of the other side.

Then we use mechanisms to protect the inference results to get private posteriors: $\mathsf{DL}(\boldsymbol{\alpha}^*)$. When we measure the distance between true posterior and protected posterior, we will have different result on l_1 norm and Hellinger distance. We believe that if the protected posterior is produced based on Hellinger distance, it will have a better accuracy than produced based on the l_1 norm.

3 BASELINE APPROACH - LAPLACE MECHANISM

Based on the posterior results from Bayesian inference, we calculate the Laplace mechanism updates on posterior and get the private results under Laplace mechanism. In the case of m dimension Dirichlet distribution, we will add $(m-1)$ i.i.d. Laplace noises $\{|Lap_1|, |Lap_2|, \cdots, |Lap_{m-1}|\}$ to the output, where $Lap_i = floor(Y)$, $Y \sim Lap(\frac{2}{\epsilon})$. The private posterior then will be $\mathsf{DL}(\alpha_1 + n_1 + |Lap_1|, \alpha_2 + n_2 + |Lap_2|, \cdots, \alpha_m + n - (Lap_1 + \cdots + Lap_{m-1}))$.

4 OUR APPROACH - EXPONENTIAL MECHANISM WITH SMOOTH SENSITIVITY

We define a new mechanism $\mathcal{M}_H(x)$ which is similar to the exponential mechanism where we use $\mathcal{R}$ as the set $\mathcal{R}_B$ of beta distributions with integer parameters summing up to $n+2$, as scoring function we use the Hellinger distance from $\mathsf{BI}(x)$, i.e. $\mathsf{H}(\mathsf{BI}(x), -)$, and we calibrate the noise to the smooth sensitivity [1]. The only difference is in the sensitivity part, since now we use the smooth sensitivity.

Definition 4.1. The mechanism $\mathcal{M}_H(x)$ outputs a candidate $r \in \mathcal{R}_B$ with probability

$$\Pr_{z\sim\mathcal{M}_H(x)}[z = r] = \frac{exp(\frac{-\epsilon \mathsf{H}(\mathsf{BI}(x), r)}{2S_\beta(x)})}{\Sigma_{r'\in R}\, exp(\frac{-\epsilon \mathsf{H}(\mathsf{BI}(x), r')}{2S_\beta(x)})},$$

where $s_\beta(x)$ is the smooth sensitivity of $\mathsf{H}(\mathsf{BI}(x), -)$, calculated by:

$$S_\beta(x) = \max(\Delta_l \mathsf{H}(\mathsf{BI}(x), -), \max_{y\neq x; y\in D^n}(\Delta_l \mathsf{H}(\mathsf{BI}(y), -) \cdot e^{-\beta d(x,y)})),$$

where d is the Hamming distance between two datasets, and $\beta = \beta(\epsilon, \delta)$ is a function of ϵ and δ.

In what follows, we will use a correspondence between the probability $\Pr_{z\sim\mathcal{M}_H(x)}[z = r]$ of every $r \in \mathcal{R}_B$ and the probability $\Pr_{z\sim\mathcal{M}_H(x)}[\mathsf{H}(\mathsf{BI}(x), z) = \mathsf{H}(\mathsf{BI}(x), r)]$ for the utility score for r. In particular, for every $r \in \mathcal{R}_B$ we have:

$$\Pr_{z\sim\mathcal{M}_H(x)}[z = r] = \frac{1}{2}\Big(\Pr_{z\sim\mathcal{M}_H(x)}[\mathsf{H}(\mathsf{BI}(x), z) = \mathsf{H}(\mathsf{BI}(x), r)]\Big)$$

To see this, it is enough to notice that: $\Pr_{z\sim\mathcal{M}_H(x)}[z = r]$ is proportional too $\mathsf{H}(\mathsf{BI}(x), r)$, i.e., $u(x, z)$. We can derive, if $u(r, x) = u(r', x)$ then $\Pr_{z\sim\mathcal{M}_H(x)}[z = r] = \Pr_{z\sim\mathcal{M}_H(x)}[z = r']$. We assume the number

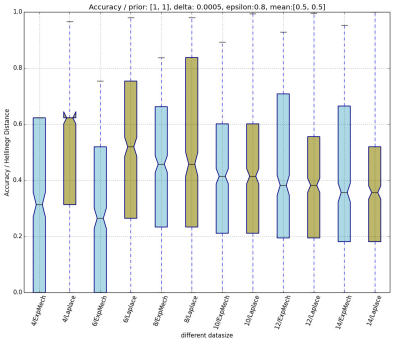
of candidates $z \in \mathcal{R}$ that satisfy $u(z,x) = u(r,x)$ is $|r|$, we have $\Pr_{z\sim\mathcal{M}_H(x)}[u(z,x) = u(r,x)] = |r| \Pr_{z\sim\mathcal{M}_H(x)}[z = r]$. Because Hellinger distance $\mathrm{H}(\mathrm{BI}(x), z)$ is axial symmetry, where the $\mathrm{BI}(x)$ is the symmetry axis. It can be infer that $|z| = 2$ for any candidates, apart from the true output, i.e., $\Pr_{z\sim\mathcal{M}_H(x)}[u(z,x) = u(r,x)] = 2 \Pr_{z\sim\mathcal{M}_H(x)}[z = r]$. This parameter can be eliminate in both sides in proof.

In our private Bayesian inference mechanism, we set the β as $\ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$.

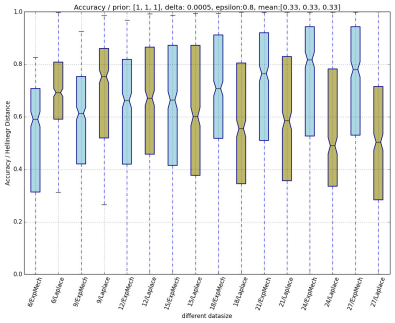
5 PRELIMINARY EXPERIMENTAL RESULTS

5.1 Accuracy Trade-off Evaluation wrt. Different Variables

In this section, we evaluate the accuracy wrt. four variables, including data size, dimension, data variance and prior distribution, and some combinations of these variables. We experiment 1000 times under each value of variables and produce 4-quantile plots for each variable. In following 4-quantile plots, the y-axis is accuracy measured by Hellinger distance, x-axis is different value of variables. The blue boxes in plots represent our exponential mechanism and the next yellow box represents the Laplace mechanism under the same setting.



(a) two dimensions with beta(1, 1) prior distribution



(b) three dimensions with DL(1, 1, 1) prior distribution

Figure 1: Accuracy measurement based on Hellinger distance wrt. different datasizes. Settings: observed data are uniformly distributed, $\epsilon = 0.8$ and $\delta = 0.0005$

5.1.1 Accuracy Evaluation wrt. Datasize. In Fig. 1, both of the two plots show that when data size go larger, accuracy of our exponential mechanism are decreasing. In Fig. 1(a), when the data size is smaller than 12, we can beta Laplace mechanism but fail when data size larger than or equal to 12. Same as in Fig. 1(b), we can beat Laplace mechanism when data size is smaller than 15 and fail otherwise.

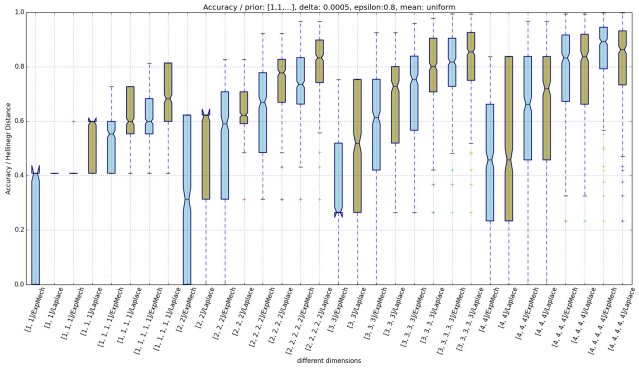


Figure 2: Accuracy measurement based on Hellinger distance wrt. different dimensions and data size. Settings: observed data are uniformly distributed, $\epsilon = 0.8$ and $\delta = 0.0005$, prior distributions are all 1 in every dimension

5.1.2 Accuracy Evaluation wrt. Dimensions. In Fig. 2, x-axis are observed data sets of different size and dimensions. The plot shows that dimensions have similar influence on our exponential mechanism and the Laplace mechanism. Accuracy of two mechanisms both decrease when dimensions go larger. We will be beat by Laplace mechanism when data size increase but will not be affected when dimensions increase. In other words, dimension has little influence on whether we will beat Laplace mechanism.

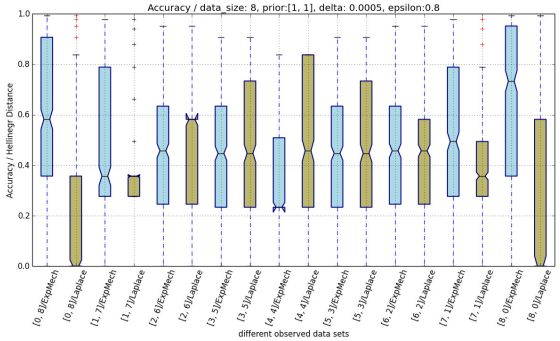


Figure 3: Accuracy measurement based on Hellinger distance wrt. different data variance. Settings: $\epsilon = 0.8$ and $\delta = 0.0005$, prior distributions are all 1 in every dimension

5.1.3 Accuracy Evaluation wrt. Data variance. In Fig. 3, x-axis are observed data sets of different variances (or means). We study

this variable under two-dimension beta distribution in order to be concise. It shows that our mechanism's accuracy is better when data variance go smaller, meanwhile Laplace mechanism go worse. We will beat Laplace mechanism when observed data are more uniformly.

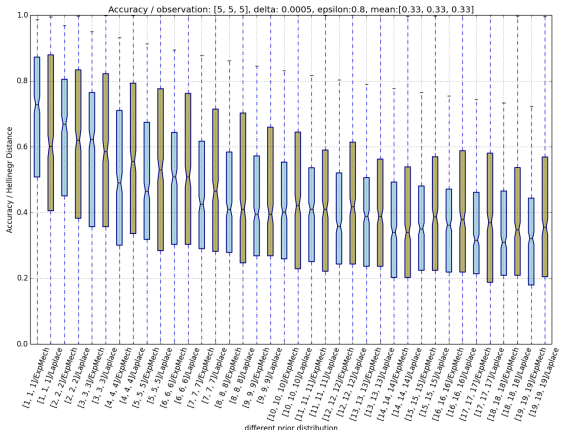


Figure 4: Accuracy measurement based on Hellinger distance wrt. different prior distribution. Settings: $\epsilon = 0.8$ **and** $\delta = 0.0005$**, observed data set is:** $[5, 5, 5]$

5.1.4 Accuracy Evaluation wrt. Prior Distribution. In Fig. 4, we study this variable under setting that observed data set is $[5, 5, 5]$ because in Fig. 1 Laplace mechanism beat us when data size is 15 and uniformly distributed. The plot shows that in the beginning we cannot beat Laplace but when prior distribution grow larger, we perform better and better and beat Laplace mechanism finally.

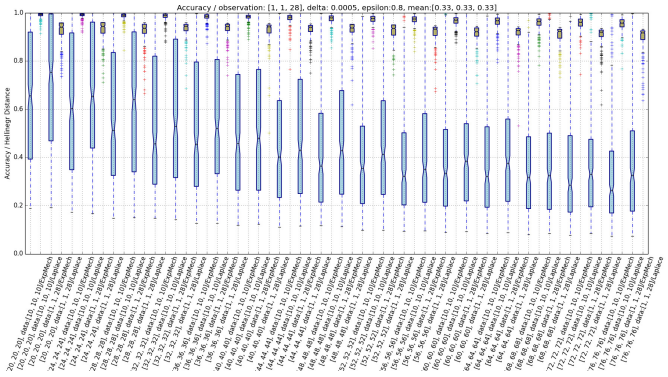


Figure 5: Accuracy measurement based on Hellinger distance wrt. different prior distribution and data variances. Settings: $\epsilon = 0.8$ **and** $\delta = 0.0005$**, observed data sets are** $[10, 10, 10]$ **and** $[1, 1, 28]$ **and prior distributions are range from** $[20, 20, 20]$ **to** $[76, 76, 76]$

5.1.5 Accuracy Evaluation wrt. Prior Distribution and Data Variance. Here, we change the prior distribution and data variance in the same time. As shown in Fig. 5, our exponential mechanism do better in uniform data set than in edging data set while Laplace mechanism on the contrary. Moreover, our mechanism is improving continuously and significantly as prior distribution increasing while Laplace mechanism isn't.

6 CONCLUSIONS

We can obtain some preliminary conclusions:

(1) We can beat Laplace mechanism when data size is small.
(2) We will beat Laplace mechanism when observed data are more uniformly.
(3) When prior distribution grow larger, we perform better and better and beat Laplace mechanism finally.

In consequence, we have a better accuracy in small data size, in larger prior and in uniformly data.

REFERENCES

[1] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2007. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. ACM, 75–84.