Notes of DP - Bayesian Inference

1 Bayesian Inference Based on Dirichlet-Bernoulli Distribution

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 $\mathsf{beta}(\alpha, \beta)$. 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.

2 Algorithm Setting up

For now, we already have a prior distribution $\pi(\boldsymbol{\xi})$, an observed data set x.

2.1 Exponential Mechanism with Global Sensitivity

In exponential mechanism, candidate set $\mathcal{R}$ can be obtained by enumerating $y \in \mathcal{X}^n$, i.e.

$$\mathcal{R} = \{\mathsf{BI}(y) \mid y \in \mathcal{X}^n\}.$$

Hellinger distance H is used here to score these candidates. The utility function:

$$u(x, r) = -\mathsf{H}(\mathsf{BI}(x), r); r \in \mathcal{R}. \tag{1}$$

Exponential mechanism with global sensitivity selects and outputs a candidate $r \in \mathcal{R}$ with probability proportional to $exp(\frac{\epsilon u(x,r)}{2\Delta_g u})$:

$$P[r] = \frac{exp(\frac{\epsilon u(x,r)}{2\Delta_g u})}{\Sigma_{r' \in \mathcal{R}}\ exp(\frac{\epsilon u(x,r')}{2\Delta_g u})},$$

where global sensitivity is calculated by:

$$\Delta_g u = \max_{\{|x',y'| \leqslant 1; x',y' \in \mathcal{X}^n\}} \max_{\{r \in \mathcal{R}\}} |\mathsf{H}(\mathsf{BI}(x'), r) - \mathsf{H}(\mathsf{BI}(y'), r)|$$

The basic exponential mechanism is $\epsilon-$differential privacy[1].

2.2 Exponential Mechanism with Local Sensitivity

Exponential mechanism with local sensitivity share the same candidate set and utility function as it with global sensitivity. This outputs a candidate $r \in \mathcal{R}$ with probability proportional to $exp(\frac{\epsilon u(x,r)}{2\Delta_l u})$:

$$P[r] = \frac{exp(\frac{\epsilon u(x,r)}{2\Delta_l u})}{\Sigma_{r' \in \mathcal{R}}\ exp(\frac{\epsilon u(x,r')}{2\Delta_l u})},$$

where local sensitivity is calculated by:

$$\Delta_l u(x) = \max_{\{|x,y'| \leqslant 1; y' \in \mathcal{X}^n\}} \max_{\{r \in \mathcal{R}\}} .\mathsf{H}(\mathsf{BI}(x), r) - \mathsf{H}(\mathsf{BI}(y'), r)|$$

The exponential mechanism with local sensitivity is non differential privacy[1].

2.3 Exponential Mechanism with Smooth Sensitivity

2.3.1 Algorithm Setting up

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 [2]. The only difference is in the sensitivity part, since now we use the smooth sensitivity.

Definition 2.1. The mechanism $\mathcal{M}_H(x)$ outputs a candidate $r \in \mathcal{R}_B$ with probability

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

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

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

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 $\underset{z \sim \mathcal{M}_H(x)}{Pr}[z = r]$ of every $r \in \mathcal{R}_B$ and the probability $\underset{z \sim \mathcal{M}_H(x)}{Pr}[\mathsf{H}(\mathsf{BI}(x), z) = \mathsf{H}(\mathsf{BI}(x), r)]$ for the utility score for r. In particular, for every $r \in \mathcal{R}_B$ we have:

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

To see this, it is enough to notice that: $\underset{z \sim \mathcal{M}_H(x)}{Pr}[z = r]$ 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 $\underset{z \sim \mathcal{M}_H(x)}{Pr}[z = r] = \underset{z \sim \mathcal{M}_H(x)}{Pr}[z = r']$. We assume the number of candidates $z \in \mathcal{R}$ that satisfy $u(z, x) = u(r, x)$ is $|r|$, we have $\underset{z \sim \mathcal{M}_H(x)}{Pr}[u(z, x) = u(r, x)] = |r| \underset{z \sim \mathcal{M}_H(x)}{Pr}[z = r]$. Because Hellinger distance $\mathsf{H}(\mathsf{BI}(x), z)$ is axial symmetry, where the $\mathsf{BI}(x)$ is the symmetry axis. It can be infer that $|z| = 2$ for any candidates, apart from the true output, i.e., $\underset{z \sim \mathcal{M}_H(x)}{Pr}[u(z, x) = u(r, x)] = 2 \underset{z \sim \mathcal{M}_H(x)}{Pr}[z = r]$. This parameter can be eliminate in both sides in proof.

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

2.3.2 Sliding Property of Exponential Mechanism

Lemma 2.1. *Consider the exponential mechanism* $\mathcal{M}_E^S(x, u, \mathcal{R})$ *calibrated on the smooth sensitivity. Let* $\lambda = f(\epsilon, \delta)$, $\epsilon \geqslant 0$ *and* $|\delta| < 1$. *Then, the following* sliding property *holds:*

$$\Pr_{r\sim\mathcal{M}_H(x)}[u(r,x) = \hat{s}] \leqslant e^{\frac{\epsilon}{2}} \Pr_{r\sim\mathcal{M}_H(x)}[u(r,x) = (\Delta + \hat{s})] + \frac{\delta}{2},$$

Proof. We denote the normalizer of the probability mass in $\mathcal{M}_H(x)$: $\sum_{r'\in\mathcal{R}} exp(\frac{\epsilon u(r',x)}{2S(x)})$ as $NL(x)$:

$$\begin{aligned}
LHS = \Pr_{r\sim\mathcal{M}_H(x)}[u(r,x) = \hat{s}] &= \frac{exp(\frac{\epsilon\hat{s}}{2S(x)})}{NL(x)} \\
&= \frac{exp(\frac{\epsilon(\hat{s}+\Delta-\Delta)}{2S(x)})}{NL(x)} \\
&= \frac{exp(\frac{\epsilon(\hat{s}+\Delta)}{2S(x)} + \frac{-\epsilon\Delta}{2S(x)})}{NL(x)} \\
&= \frac{exp(\frac{\epsilon(\hat{s}+\Delta)}{2S(x)})}{NL(x)} \cdot e^{\frac{-\epsilon\Delta}{2S(x)}}).
\end{aligned}$$

By bounding the $\Delta \geqslant -S(x)$, we can get:

$$\begin{aligned}
\frac{exp(\frac{\epsilon(\hat{s}+\Delta)}{2S(x)})}{NL(x)} \cdot e^{\frac{-\epsilon\Delta}{2S(x)}} &\leqslant \frac{exp(\frac{\epsilon(\hat{s}+\Delta)}{2S(x)})}{NL(x)} \cdot e^{\frac{\epsilon}{2}} \\
&= e^{\frac{\epsilon}{2}} \Pr_{z\sim\mathcal{M}_H(x)}[u(r,x) = (\Delta + \hat{s})] \leqslant RHS
\end{aligned}$$

□

2.3.3 Dilation Property of Exponential Mechanism

Lemma 2.2. *for any exponential mechanism* $\mathcal{M}_H(x)$, $\lambda < |\beta|$, ϵ, $|\delta| < 1$ *and* $\beta \leqslant \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$, *the dilation property holds:*

$$\Pr_{r\sim\mathcal{M}_H(x)}[u(r,x) = z] \leqslant e^{\frac{\epsilon}{2}} \Pr_{r\sim\mathcal{M}_H(x)}[u(r,x) = e^{\lambda}z] + \frac{\delta}{2},$$

where the sensitivity in mechanism is still smooth sensitivity as above.

Proof. The sensitivity is always greater than 0, and our utility function $-\mathsf{H}(\mathsf{BI}(x), r)$ is smaller than zero, i.e., $u(r,x) \leqslant 0$, we need to consider two cases where $\lambda < 0$, and $\lambda > 0$:

We set the $h(z) = Pr[u(\mathcal{M}_H(x)) = z] = 2\frac{exp(\frac{\epsilon z}{2S(x)})}{NL(x)}$.

We first consider $\lambda < 0$. In this case, $1 < e^{\lambda}$, so the ratio $\frac{h(z)}{h(e^{\lambda}z)} = \frac{exp(\frac{\epsilon z}{2S(x)})}{exp(\frac{\epsilon(z\cdot e^{\lambda})}{2S(x)})}$ is at most $\frac{\epsilon}{2}$.

Next, we proof the dilation property for $\lambda > 0$, The ratio of $\frac{h(z)}{h(e^{\lambda}z)}$ is $\exp(\frac{\epsilon}{2} \cdot \frac{u(\mathcal{M}_H(x))(1-e^{\lambda})}{S(x)})$. Consider the event $G = \{\mathcal{M}_H(x) : u(\mathcal{M}_H(x)) \leqslant \frac{S(x)}{(1-e^{\lambda})}\}$. Under this event, the log-ratio above is at most $\frac{\epsilon}{2}$. The probability of G under density $h(z)$ is $1 - \frac{\delta}{2}$. Thus, the probability of a given event z is at most $Pr[z \cap G] + Pr[\overline{G}] \leqslant e^{\frac{\epsilon}{2}} Pr[e^{\lambda}z \cap G] + \frac{\delta}{2} \leqslant e^{\frac{\epsilon}{2}} Pr[e^{\lambda}z] + \frac{\delta}{2}$.

Detail proof:

- $\lambda < 0$

  The left hand side will always be smaller than 0 and the right hand side greater than 0. This will always holds, i.e.

- $\lambda > 0$

  Because $\hat{s} = u(r)$ where $r \sim \mathcal{M}_H(x)$, we can substitute $\hat{s}$ with $u(\mathcal{M}_H(x))$. Then, what we need to proof under the case $\lambda > 0$ is:

$$u(\mathcal{M}_H(x)) \leqslant \frac{S(x)}{(1-e^{\lambda})}$$

  By applying the accuracy property of exponential mechanism, we bound the probability that the equation holds with probability:

$$Pr[u(\mathcal{M}_H(x)) \leqslant \frac{S(x)}{(1-e^{\lambda})}] \leqslant \frac{|\mathcal{R}|exp(\frac{\epsilon S(x)}{(1-e^{\lambda})}/2S(x))}{|\mathcal{R}_{OPT}|exp(\epsilon OPT_{u(x)}/2S(x))}$$

  In our Bayesian Inference mechanism, the size of the candidate set $\mathcal{R}$ is equal to the size of observed data set plus 1, i.e., $n+1$, and $OPT_{u(x)} = 0$, then we have:

$$\begin{aligned} Pr[u(\mathcal{M}_H(x)) \leqslant \frac{S(x)}{(1-e^{\lambda})}] &= (n+1)exp(\frac{\epsilon S(x)}{(1-e^{\lambda})}/2S(x)) \\ &= (n+1)exp(\frac{\epsilon}{2(1-e^{\lambda})}) \end{aligned}$$

  When we set $\lambda \leqslant \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$, it is easily to derive that $Pr[u(\mathcal{M}_H(x)) \leqslant \frac{S(x)}{(1-e^{\lambda})}] \leqslant \frac{\delta}{2}$.

□

2.3.4 Privacy Analysis

Lemma 2.3. *$\mathcal{M}_H$ is (ϵ, δ)-differential privacy.*

Proof. of Lemma 2.3: For all neighboring $x, y \in D^n$ and all sets $\mathcal{S}$, we need to show that:

$$\underset{z\sim\mathcal{M}_H(x)}{Pr}[z \in \mathcal{S}] \leqslant e^{\epsilon} \underset{z\sim\mathcal{M}_H(y)}{Pr}[z \in \mathcal{S}] + \delta.$$

Given that $2\Big(\underset{z\sim\mathcal{M}_H(x)}{Pr}[z \in \mathcal{S}]\Big) = \underset{z\sim\mathcal{M}_H(x)}{Pr}[u(x,z) \in \mathcal{U}]$, let $\mathcal{U}_1 = \frac{u(y,z)-u(x,z)}{S(x)}$, $\mathcal{U}_2 = \mathcal{U} + \mathcal{U}_1$ and $\mathcal{U}_3 = \mathcal{U}_2 \cdot \frac{S(x)}{S(y)} \cdot \ln(\frac{NL(x)}{NL(y)})$. Then,

$$\begin{aligned} 2\Big(\underset{z\sim\mathcal{M}_H(x)}{Pr}[z \in \mathcal{S}]\Big) &= \underset{z\sim\mathcal{M}_H(x)}{Pr}[u(x,z) \in \mathcal{U}] \\ &\leqslant e^{\epsilon/2} \cdot \underset{z\sim\mathcal{M}_H(x)}{Pr}[u(x,z) \in \mathcal{U}_2] \\ &\leqslant e^{\epsilon} \cdot \underset{z\sim\mathcal{M}_H(x)}{Pr}[u(x,z) \in \mathcal{U}_3] + e^{\epsilon/2} \cdot \frac{\delta'}{2} \\ &= e^{\epsilon} \cdot \underset{z\sim\mathcal{M}_H(y)}{Pr}[u(y,z) \in \mathcal{U}] + \delta = 2\Big(e^{\epsilon} \cdot \underset{z\sim\mathcal{M}_H(x)}{Pr}[z \in \mathcal{S}]\Big) + \delta \end{aligned}$$

The first inequality holds by the sliding property, since the $\mathcal{U}_1 \geqslant -S(x)$. The second inequality holds by the dilation property, since $\frac{S(x)}{S(y)} \cdot \ln(\frac{NL(x)}{NL(y)}) \leqslant 1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})}$.

□

3 Accuracy Analysis

3.1 Laplace Mechanism

Fixing a data set x, we already had the accuracy bound based on the l_1 norm as:

$$Pr[||Lap(\mathsf{BI}(x)) - \mathsf{BI}(x)||_1 \geqslant |x| \ln(\frac{m}{\gamma})\frac{2}{\epsilon}] \leqslant \gamma,$$

where m is the order of the Dirichlet distribution. Then we have the accuracy bound based on H distance.

$$\begin{aligned} Pr[||Lap(\mathsf{BI}(x)) - \mathsf{BI}(x)||_1 \geqslant m \ln(\frac{m}{\gamma})\frac{2}{\epsilon}] &\leqslant Pr[m||Lap(\mathsf{BI}(x)) - \mathsf{BI}(x)||_\infty \geqslant m \ln(\frac{m}{\gamma})\frac{2}{\epsilon}] \\ &= Pr[||Lap(\mathsf{BI}(x)) - \mathsf{BI}(x)||_\infty \geqslant \ln(\frac{m}{\gamma})\frac{2}{\epsilon}] \\ &= \gamma \end{aligned}$$

3.2 Exponential Mechanism with Global Sensitivity

3.3 Exponential Mechanism with Local Sensitivity

3.4 Exponential Mechanism with Smooth Sensitivity

4 Experimental Evaluations

4.1 Computation Efficiency

The formula for computing the local sensitivity is presented in Sec. 2.2: $\max_{\{|x,y'| \leqslant 1; y' \in \mathcal{X}^n\}} \max_{\{r \in R\}} \{\mathsf{H}(\mathsf{BI}(x), r) - \mathsf{H}(\mathsf{BI}(y'), r)|\}$ can be reduced to $\max_{\{|x,y'| \leqslant 1; y' \in \mathcal{X}^n\}} \mathsf{H}(\mathsf{BI}(x), \mathsf{BI}(y'))|$ by applying the distance triangle property. i.e., the maximum value over $\max_{r \in R}$ always happen when $r = \mathsf{BI}(x)$ itself, where $\Delta_l u(x) = \max_{\{|x,y'| \leqslant 1; y' \in \mathcal{X}^n\}} \{\mathsf{H}(\mathsf{BI}(x), \mathsf{BI}(x)) - \mathsf{H}(\mathsf{BI}(y'), \mathsf{BI}(x))|\} = \max_{\{|x,y'| \leqslant 1; y' \in \mathcal{X}^n\}} \{\mathsf{H}(\mathsf{BI}(y'), \mathsf{BI}(x))|\}$. We also have some experiments for validating our proposal as in Fig. 4.1, where we calculate the $\max_{\{|x,y'| \leqslant 1; y' \in \mathcal{X}^n\}}$ value for every candidate $r \in R$. It is shown that maximum value taken when $r = \mathsf{BI}(x)$.

4.2 Accuracy Study Based on Hellinger Distance

We study the accuracy property of the exponential mechanism with three kind of sensitivity, as well as the Laplace mechanism for comparison. As in Fig. 2, the two figures shown the accuracy of four algorithms under same configuration. The accuracy is measured by Hellinger distance between output r^* and the true inference result $\mathsf{BI}(x)$, $\mathsf{H}(r^*, \mathsf{BI}(x))$.

Then, we extended the beta to the Bernoulli distribution, got the result in Fig. 3

4.3 Accuracy Study Based on L_1 Norm

For comparison, we study the accuracy of these four mechanisms based on l_1 norm, with the Dirichlet distribution $\mathsf{DL}(7, 4, 5)$ and data size 150, shown in Fig. 4.

4.4 Accuracy Study in Theory

We studied the accuracy of our exponential mechanism and Laplace mechanism in theory. By computing the discrete probability of outputting each candidate in two mechanisms separately. We count the probabilities wrt. the steps from correct answer. For example, in the case where the correct posterior distribution is $\mathsf{beta}(5, 5)$, the candidate $\mathsf{beta}(4, 6)$ and $\mathsf{beta}(6, 4)$ are of 1 steps from $\mathsf{beta}(5, 5)$; when correct posterior distribution is

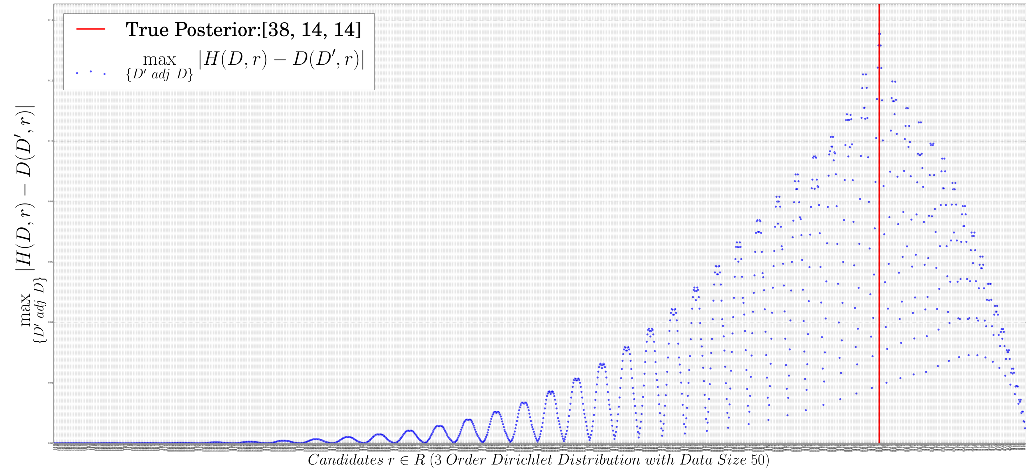


Figure 1: Experimental Results for Finding the Local Sensitivity Efficiently

$\mathsf{DL}(5,5,5)$, the candidate $\mathsf{DL}(4,5,3)$, $\mathsf{DL}(4,3,5)$, $\mathsf{DL}(5,4,3)$, $\mathsf{DL}(5,3,4)$, $\mathsf{DL}(3,5,4)$ and $\mathsf{DL}(5,5,5)$ are all of 1 step from $\mathsf{DL}(5,5,5)$. Under the Hellinger distance measurement, if two candidates are of the same steps from correct answer, they also have the same Hellinger distance from correct answer. i.e for every $\mathsf{H}(\mathsf{BI}(x), z) = c$, it will have a corresponding steps $step(\mathsf{H}(\mathsf{BI}(x), z) = c) = k$.

So in order to be more concise and comparable to the experiment results in Sec. 4.2, we sum up the probabilities values of candidates whose steps (i.e., the hellinger distance) from the correct answer are the same:

$$\underset{z \sim \mathcal{M}_H(x)}{Pr}[\mathsf{H}(\mathsf{BI}(x), z) = c] = \sum_{\{z \mid \mathsf{H}(\mathsf{BI}(x), z) = c\}} \frac{exp(\frac{-\epsilon c}{S(x)})}{NL_x}$$

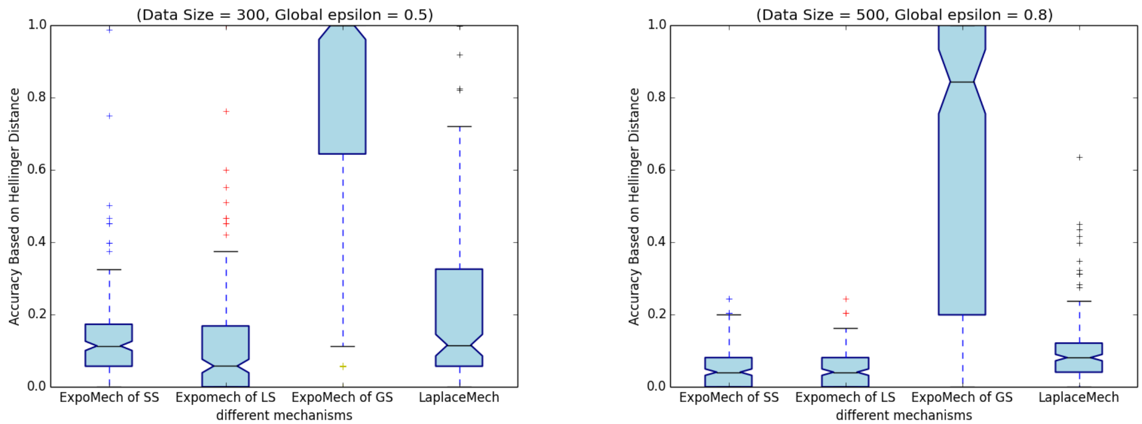
Then we get groups of discrete probabilities wrt. the steps (i.e., Hellinegr distance) from the correct answer, and plot them. Under different settings, we got same results as in Fig. 5, 6 and 7. The x-axis is the hellinger distance from correct answer, i.e. steps from correct answer. Since they have the same meaning, here we just use the hellinger distance to label the x-axis. The y-axis is the probability of output candidates with corresponding steps from correct answer.

As shown in Fig. 5, our exponential mechanism is better than Laplace, both in experiments and in theory. In Fig. 5 (a), the experiment results shown that the average accuracy of our mechanism is better than Laplace mechanism. In Our exponential mechanism can output good answer with higher probability and output bad answer with lower probability.

In Fig. 6, we plotted these discrete probabilities under the case (x=(20,20,20)), I think this plot shows the laplace mech performs better than ExpMech to some extent but not significant.

In Fig. 7, we plotted these discrete probabilities under the case ($x = (20, 20, 20, 20)$). This plot can show clearly that the Laplace Mech are better than Exp Mech when the size is larger and order is higher. The Laplace can output good answers with much higher probabilities, even though their performances on bad answers are not so obviously.

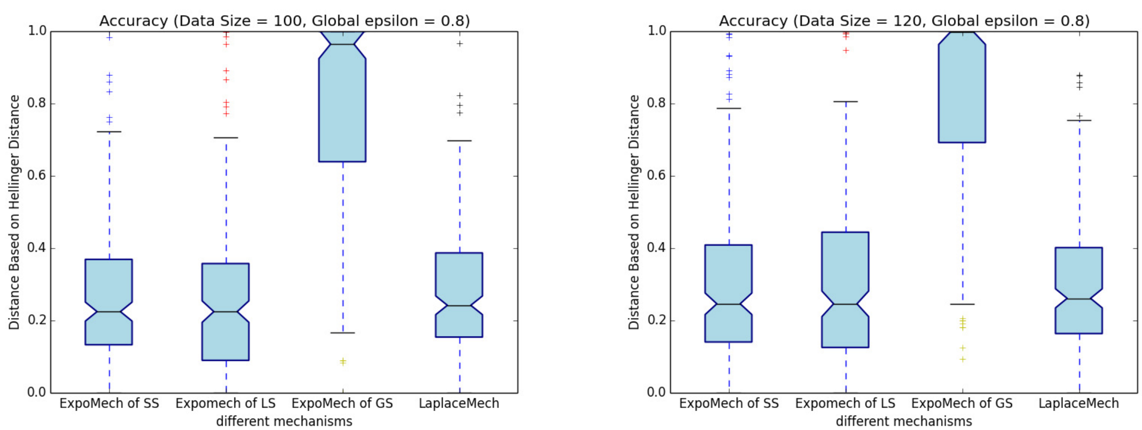
We also consider an edge case, In Fig. 8, we plotted these discrete probabilities under the case ($x = (1, 1, 1, 77)$). This plot can show clearly that the Laplace Mech are better than Exp Mech when the size is larger and order is higher. The Laplace can output good answers with much higher probabilities, even though their performances on bad answers are not so obviously.



(a) Data size $n = 500$ with global $\epsilon = 0.5$

(b) Data size $n = 300$ with global $\epsilon = 0.5$

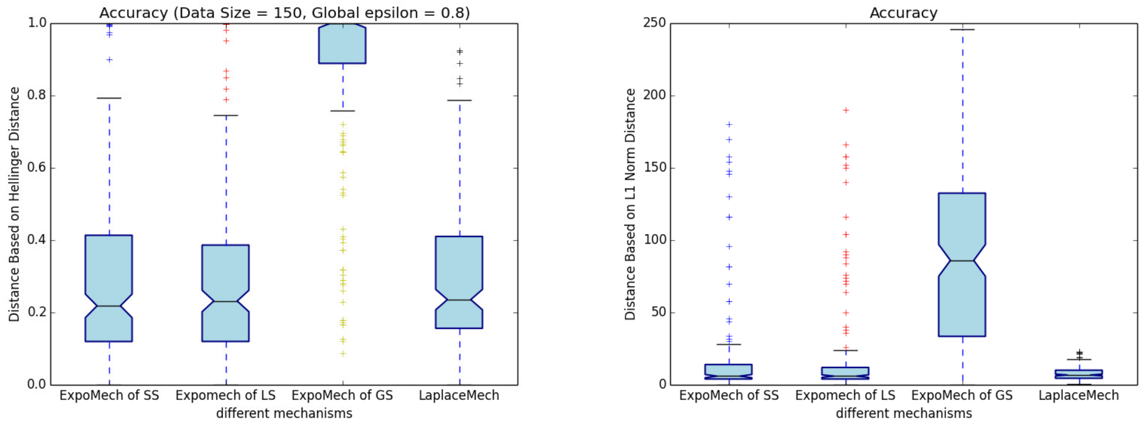
Figure 2: The experimental results of accuracy of algorithms with Beta prior distribution $\mathsf{beta}(7, 4)$ based on Hellinger distance



(a) Data size $n = 100$, the exponential mechanism with global sensitivity 0.239992747797 is 0.8 -DP, with local sensitivity 0.08 is Non-Private and with 0.0699407108115 - bound smooth sensitivity 0.09 is (0.8,0.8)-DP

(b) Data size $n = 120$, the exponential mechanism with global sensitivity 0.239992747797 is 0.8 -DP, with local sensitivity 0.0945 is Non-Private, with 0.0677791100173 - bound smooth sensitivity 0.096 is (0.8,0.8)-DP

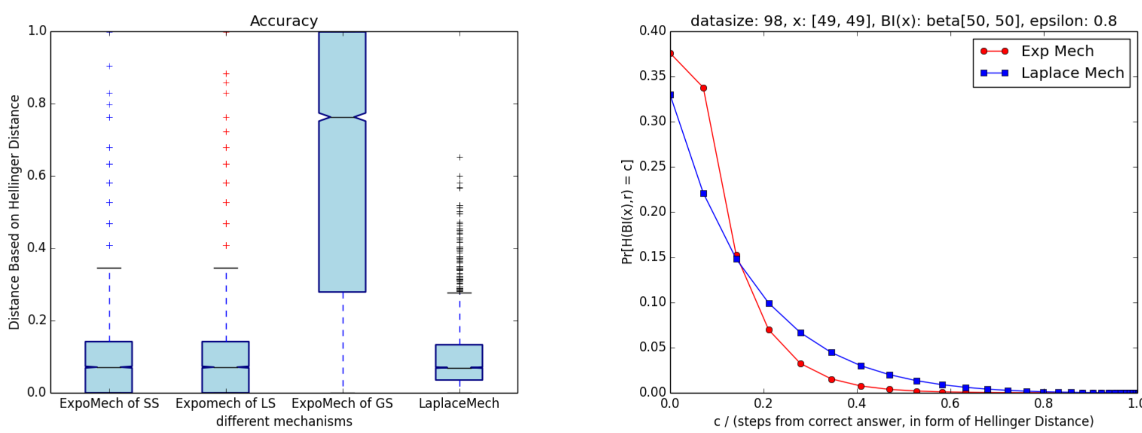
Figure 3: The experimental results of accuracy of algorithms with Dirichlet prior distribution $\mathsf{DL}(7, 4, 5)$ based on Hellinger distance



(a) accuracy measurement based on Hellinger distance

(b) accuracy measurement based on l_1 norm

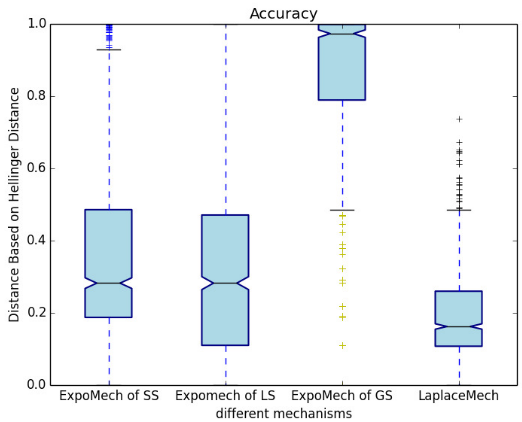
Figure 4: The experimental results of accuracy of algorithms with Dirichlet prior distribution $\mathsf{DL}(7, 4, 5)$ based on Hellinger distance and l_1 norm, where data size $n = 150$, the exponential mechanism with global sensitivity 0.239992747797 is 0.8 -DP, with local sensitivity 0.0945 is Non-Private, with 0.0677791100173 -bound smooth sensitivity 0.096 is (0.8,0.8)-DP



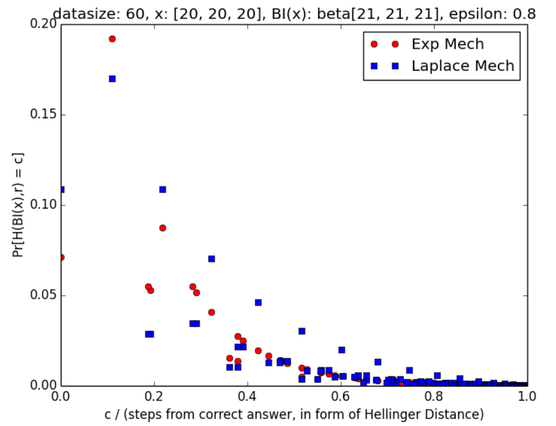
(a) accuracy measurement based on Hellinger distance

(b) discrete probabilities wrt. hellinger distance

Figure 5: Settings: Dirichlet prior distribution $\mathsf{DL}(1, 1)$, observed data $x = (49, 49)$,

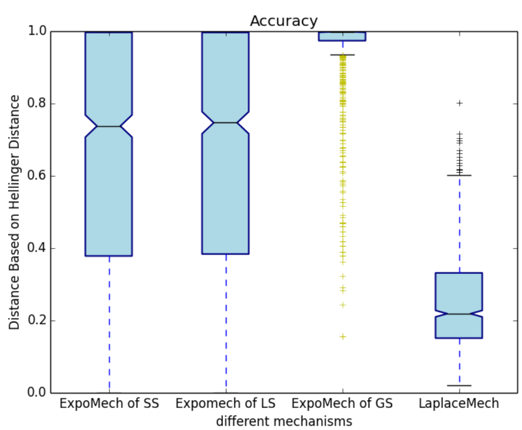


(a) accuracy measurement based on Hellinger distance

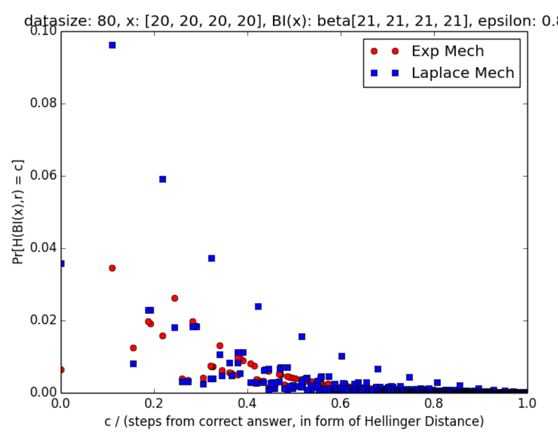


(b) discrete probabilities wrt. hellinger distance

Figure 6: Settings: Dirichlet prior distribution $\mathsf{DL}(1, 1, 1)$, observed data $x = (20, 20, 20)$,

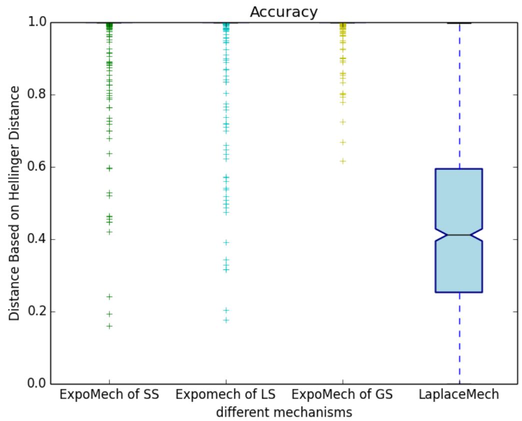


(a) accuracy measurement based on Hellinger distance

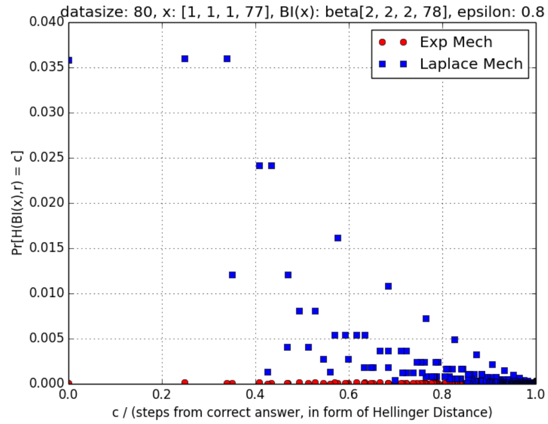


(b) discrete probabilities wrt. hellinger distance

Figure 7: Settings: Dirichlet prior distribution $\mathsf{DL}(1, 1, 1, 1)$, observed data $x = (20, 20, 20, 20)$,



(a) accuracy measurement based on Hellinger distance



(b) discrete probabilities wrt. hellinger distance

Figure 8: Settings: Dirichlet prior distribution $\mathsf{DL}(1,1,1,1)$, observed data $x = (1,1,1,77)$,

References

[1] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

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