# 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 $\mathsf{H}$ 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

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

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

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

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

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 $\mathsf{H}(\mathsf{BI}(x), z)$ 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., $\Pr_{z \sim \mathcal{M}_H(x)}[u(z, x) = u(r, x)] = 2 \Pr_{z \sim \mathcal{M}_H(x)}[z = r]$. This parameter can be eliminate in both sides in proof.

In our private Bayesian inference mechanism, we set the $\beta$ 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 $\mathcal{M}_H(x)$, $\lambda < |\beta|$, $\epsilon$, $|\delta| < 1$ 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 $(\epsilon, \delta)$-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:

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

Given that $2\Big(\Pr_{z\sim\mathcal{M}_H(x)}[z \in \mathcal{S}]\Big) = \Pr_{z\sim\mathcal{M}_H(x)}[u(x,z) \in \mathcal{U}]$, 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,

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

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 $\mathsf{H}$ 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

In this section, we do some experiments in order to study the accuracy property of these mechanisms, including the exponential mechanism with three kind of sensitivity and the Laplace mechanism.

### 4.2.1 Accuracy Study Based on Hellinger Distance

We first analyzed the accuracy based on the Hellinger distance. By repeating the experiments for 10000 times and plotting the accuracy after each execution, we got two groups of results under different parameters setting as in Fig. 2 and Fig. 3. X-axis is labeled with different mechanisms. We took four mechanisms in our experiments: our newly designed exponential mechanism with smooth sensitivity named "ExpoMech of SS", exponential mechanism with local sensitivity named " ExpoMech of LS", exponential mechanism with global sensitivity named " ExpoMech of GS" and Laplace mechanism named "LaplaceMech". Y-axis is the accuracy measured by Hellinger distance between output $r$ of each execution and the correct inference result $\mathsf{BI}(x)$, $\mathsf{H}(r, \mathsf{BI}(x))$.

When prior distribution is $\mathsf{beta}$, under data size $n = 500$ and 300, we obtained two plots in Fig. 2. It is shown that our ExpoMech of SS is slightly better than Laplace mechanism and much more better than it with global sensitivity.

To have more experimental results, we then extended the $\mathsf{beta}$ to the $\mathsf{DL}$ distribution as well as increased the data size, plotted in Fig. 3.

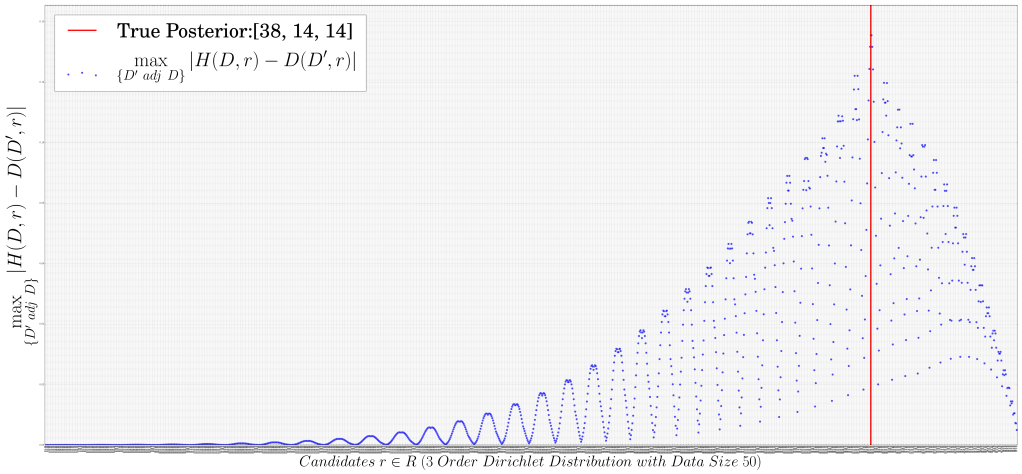



Figure 1: Experimental Results for Finding the Local Sensitivity Efficiently

In both of the two cases, the performances of our exponential mechanism with smooth sensitivity and Laplace mechanism are very close. We can obtain some preliminary conclusions:

1. Our exponential mechanism with smooth sensitivity can have similar accuracy as it with local sensitivity. Both of the two exponential mechanisms are do much better than it with global sensitivity.

2. The accuracy of our exponential mechanism is better than Laplace mechanism. But the advantages are not significant.

So we need to have a further exploration on the accuracy trade-off between Laplace mechanism and our exponential mechanism.

#### 4.2.2 Accuracy Study Based on $L_1$ Norm

In this section, we will 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 for comparison, as in Fig. 4.

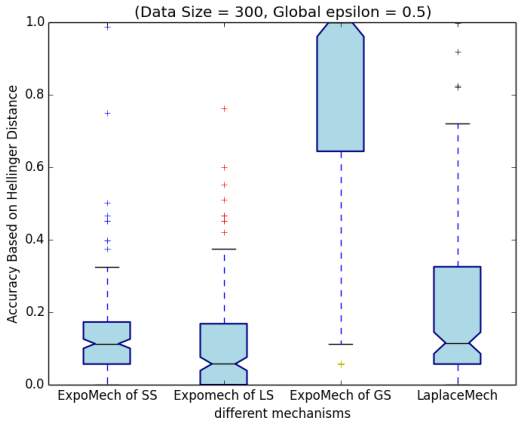
The results based on $l_1$ norm are actually very similar to results based on Hellinegr distance.

### 4.3 Accuracy Trade-off Between Laplace and Our Exponential Mechanism with Smooth Sensitivity

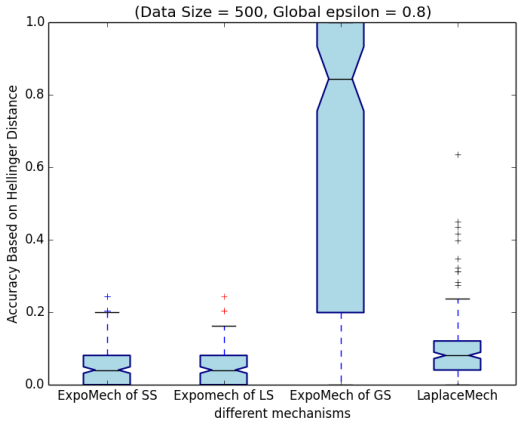
Based on accuracy study in Sec. 4.2.1 and 4.2.2, we are going to have a further study on the relationship between Laplace mechanism and our exponential mechanism. We will do analysis both in theory and experiment based on accuracy bound and probability formula, and then study on some concrete cases.

#### 4.3.1 Theory Analysis

In this part, we will do theory analysis on the two mechanisms' discrete probabilities of outputting each candidate separately. In following analysis, we will count the probabilities wrt. the steps from correct answer, in order to be more concise and comparable to the experiment results from Sec. 4.2.1. 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 $\mathsf{DL}(5, 5, 5)$, 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

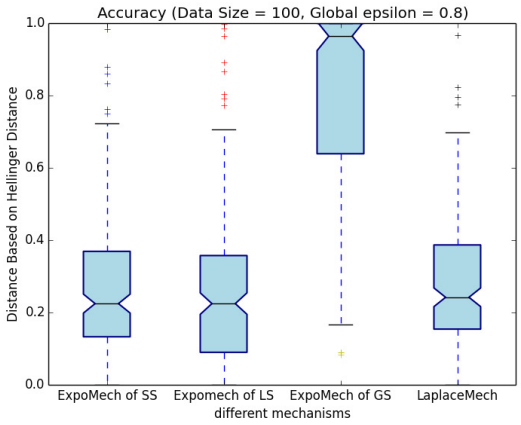



(a) Data size $n = 500$ with global $\epsilon = 0.5$

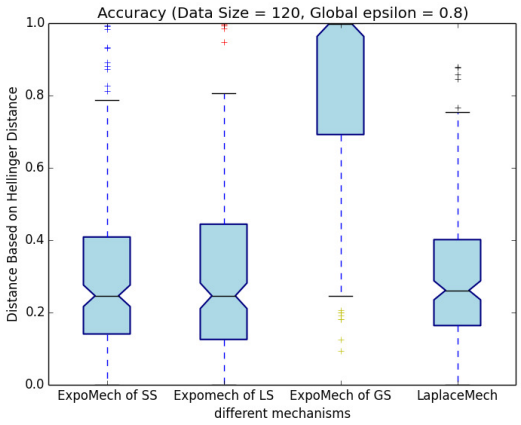



(b) Data size $n = 300$ with global $\epsilon = 0.5$

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

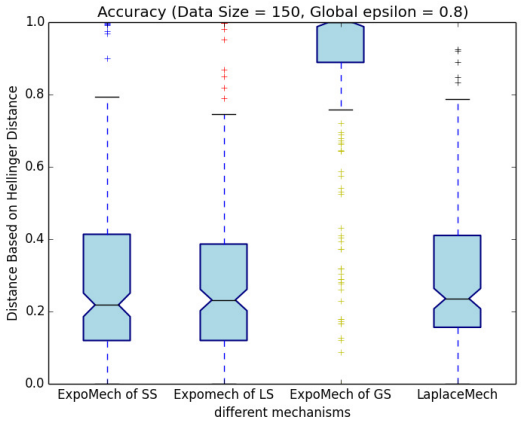



(a) Data size $n = 100$, the exponential mechanism with global sensitivity 0.239992747797 is 0.8 -DP, with local sensitivity 0.08 is Non-Private and with 0.0699407108115 - bound smooth sensitivity 0.09 is (0.8,0.8)-DP

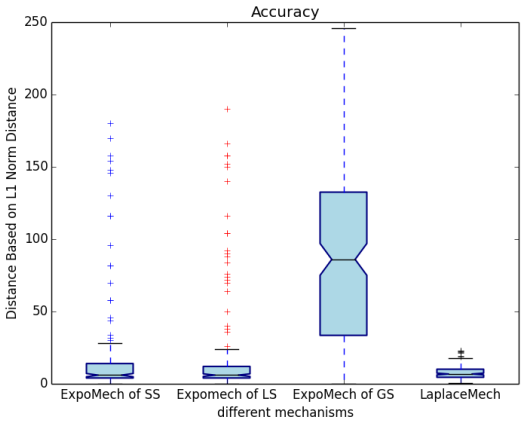



(b) Data size $n = 120$, the exponential mechanism with global sensitivity 0.239992747797 is 0.8 -DP, with local sensitivity 0.0945 is Non-Private, with 0.0677791100173 - bound smooth sensitivity 0.096 is (0.8,0.8)-DP

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

distance measurement, if 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 $step(\mathsf{H}(\mathsf{BI}(x), z) = c) = k$. This will make the results more clear and observable.

- In our exponential mechanism, the probability of outputting candidates wrt. steps (i.e., the hellinger distance) from the correct answer is calculated as:

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

  Each candidate will occupy a portion of "1" as their outputting probability. Supposing candidates within three steps from the correct answer are good answers, the portion occupied by the good answers is decreasing when the size of candidate set increasing. That's to say, the probabilities of outputting good answers are decreasing when the data size and dimension of prior distribution increasing. More specifically, the candidate set size is $\sim n^{m-1}$, which means the probabilities of outputting good answers are decreasing with speed $\sim n^{m-1}$.

- However in Laplace mechanism, the probability of producing noise has little relevance with the size of the candidate set. In $m$ dimensional Dirichlet distribution, when the correct posterior distribution is $\mathsf{DL}(\alpha_1, \alpha_2, \cdots, \alpha_m)$, we are adding Laplace noise in this way: $\mathsf{DL}(\alpha_1 + Lap_1, \alpha_2 + Lap_2, \cdots, n - (\alpha_1 + Lap_1 + \alpha_{m-1} + Lap_{m-1}))$ where $Lap_i \sim Floor(Lap(\frac{2}{\epsilon}))$ identical and independently. So we will have a list of noise $L_{noise} = \{|Lap_1|, |Lap_2|, \cdots, |Lap_{m-1}|\}$ by taking the absolute value of every noise $|Lap_i|$. In order to get the probability of one candidate, we need to have a set $S_{noise}$ obtained from $L_{noise}$ by removing the duplicate values, a list $L_{noise/0}$ by removing the value 0. Then, the probabilities of outputting one candidate in Laplace mechanism can be computed by:

$$
\begin{aligned}
Pr[(Lap_1, Lap_2, \cdots, Lap_{m-1})] &= \frac{1}{|S_{noise}|! \times 2^{|L_{noise/0}|}} \times Pr[(|Lap_1|, |Lap_2|, \cdots, |Lap_{m-1}|)] \\
&= \frac{1}{|S_{noise}|! \times 2^{|L_{noise/0}|}} \times Pr[|Lap_1| \leqslant Lap(\frac{2}{\epsilon}) < |Lap_1| + 1] \\
&\times Pr[|Lap_2| \leqslant Lap(\frac{2}{\epsilon}) < |Lap_2| + 1] \times \cdots \\
&\times Pr[|Lap_{m-1}| \leqslant Lap(\frac{2}{\epsilon}) < |Lap_{m-1}| + 1]
\end{aligned}
$$

By summing up the probability value of candidates whose steps from correct answer are the same, we can finally get the probability wrt. steps from correct answer.

From analysis above, we can see the probability of output the good answer will not changed a lot as the size of the candidate set increasing. Specifically, we can easily have the two upper bounds: $|S_{noise}| \leqslant (m-1)$ and $|L_{noise/0}| \leqslant (m-1)$. As a result, the probabilities of outputting good answers are decreasing with speed upper bounded by $(m-1)! \times 2^{m-1}$. Since $n \gg 2$ and $m$ is usually smaller than 10, it is easy to see that the decay speed $(m-1)! \times 2^{m-1}$ in Laplace mechanism is much more smaller than $n^{m-1}$ in our exponential mechanism. Moreover, when the steps are small, 0 for example, $|S_{noise}| = |L_{noise/0}| = 0$, which means the probability of correct answer will decrease very little no matter how large the candidate set is.

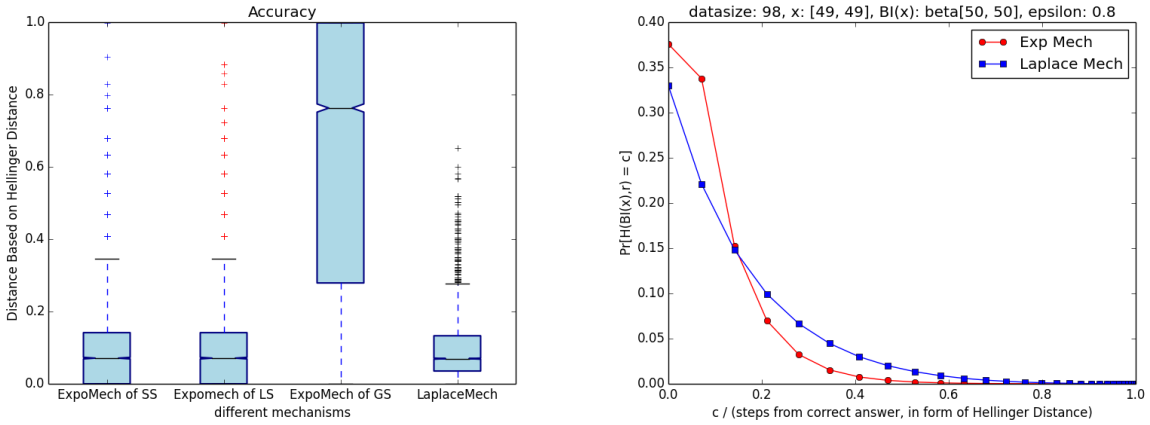
- Then, we do some concrete cases analysis.
  - when the prior is $\mathsf{beta}(1,1)$, the observed data set is $(1,1,0,0,1,1,0,0)$, it is easy to compute the posterior distribution: $\mathsf{beta}(5,5)$, the probability from the two mechanisms are:

| Hellinger Distance / steps | Our Exponential Mechanism | Laplace Mechanism |
|---|---|---|
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.83737258593]/4$ | 0.0431193490585 | 0.066561234758 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.662174391701]/3$ | 0.0785621424847 | 0.0992976939175 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.457635865026]/2$ | 0.158265808563 | 0.148134752205 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.233629480709]/1$ | 0.340809715054 | 0.220991081918 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.0]/0$ | 0.37924298484 | 0.329679953964 |

  - when the prior is $\mathsf{DL}(1,1,1)$, the observed data set is $(20,20,20)$ (suppose a black box will produce $A$, $B$, $C$ with a certain distribution, after observing this black box continuously for 60 times, we get 20 times $A$, 20 times $B$ and 20 times $C$), we can get the posterior distribution $\mathsf{DL}(21,21,21)$

| Hellinger Distance / steps | Our Exponential Mechanism | Laplace Mechanism |
|---|---|---|
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.999999984481]/\cdots$ | 0.000149705644585 | 3.05988187701e-09 |
| | $\cdots$ | |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.187421762881]/2$ | 0.0548161224677 | 0.0285774941516 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.110122822057]/1$ | 0.192227323562 | 0.170131188571 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.0]/0$ | 0.0713016293602 | 0.108688872046 |

    In this case, it is hard to tell which one is better because Laplace mechanism can output some good answers (for example, the correct answer) better than our mechanism but cannot go better with other good answers.
  - when the prior is $\mathsf{DL}(1,1,1,1)$, the observed data set is $(1,1,1,50)$ (the same meaning as above), we can get the posterior distribution $\mathsf{DL}(2,2,2,51)$. Some probabilities from the two mechanisms are listed as follows:




(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 $x = (49, 49)$, $\epsilon = 0.8$ and $\delta = 0.000005$

| Hellinger Distance / steps | Our Exponential Mechanism | Laplace Mechanism |
|---|---|---|
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.999999999998]/\cdots$ | 0.000213865923868 | 6.16388940464e-11 |
| | ... | |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.340503311163]/2$ | 0.000388935212208 | 0.0360289071389 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.249722620018]/1$ | 0.000464587461035 | 0.0360289071389 |
| $Pr[\mathsf{H}(\mathsf{BI}(x), r) = 0.0]/0$ | 0.000252512987228 | 0.0358325423325 |

This case shows that the Laplace mechanism do much better than our exponential mechanism significantly. These good answers with few steps from correct answer can be outputted with higher probabilities in Laplace mechanism than in our mechanism. Moreover, in our mechanism the bod answer and good answer have very similar outputting probabilities.

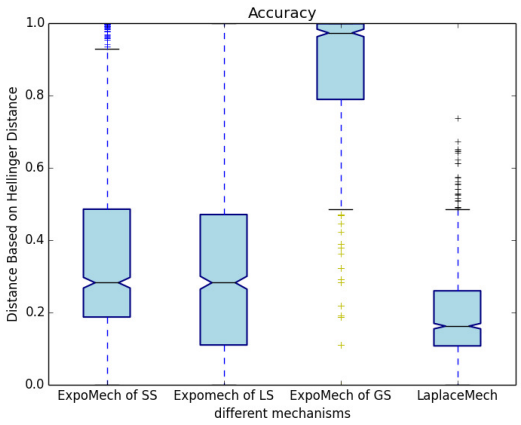
### 4.3.2 Experiment Analysis

In this part, we will do some experiment analysis to study the accuracy trade-off between our exponential mechanism and Laplace mechanism. By computing the two mechanisms' discrete probabilities wrt. steps from correct answer. Then we got groups of discrete probabilities wrt. the steps (i.e., Hellinegr distance) from the correct answer, and plotted them. Under different settings, we got same results as in Fig. 5, 6 and 7.

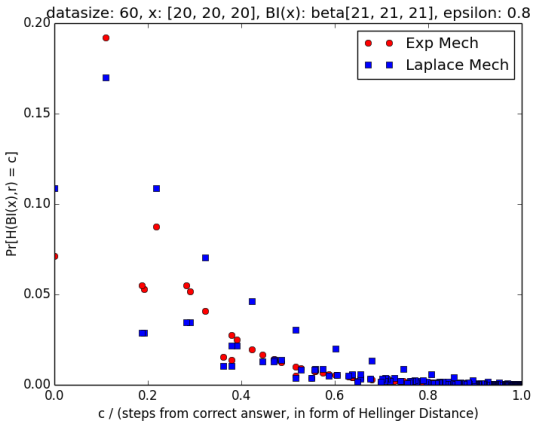
Subfigs (a) are from Sec. 4.2.1 shows the average accuracy of these mechanisms by executing them for 10000 times. x-axis is labeled with mechanisms and y-axis is the accuracy of every execution outputs measured by Hellinegr distance. Subfigs (b) are concrete probabilities of Laplace and our exponential mechanisms' wrt. the steps. 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 (i.e. Hellinger distance) from correct answer.

When the data size and the dimension of the prior distribution is small, our exponential mechanism can do better than Laplace, both in experiments and in theory, as shown in Fig. 5. In Fig. 5 (a), the experiment results show that the average Hellinger distance of our mechanism's outputs are smaller than Laplace mechanism, i.e., the average accuracy of our mechanism is higher than Laplace mechanism. In the mean time, the theory calculation results in Fig. 5 (b) also shows our exponential mechanism can output good answer with higher probability and output bad answer with lower probability as well.

In Fig. 6, we plotted the results under the case ($x = (20, 20, 20)$). SubFig. 6(a) shows that the average accuracy of Laplace mechanism is slightly better than our mechanism. Meanwhile, the SubFig. 6(b) shows the Laplace mechanism have higher probabilities on some good answers but lower on some other good answers. The advantage is not significant. So, we increase the data size and the dimension to have the next case.

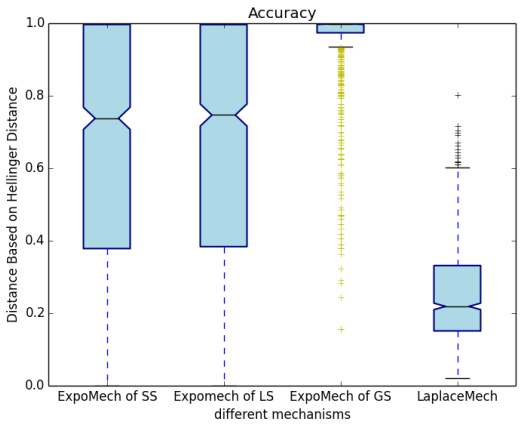



(a) accuracy measurement based on Hellinger distance

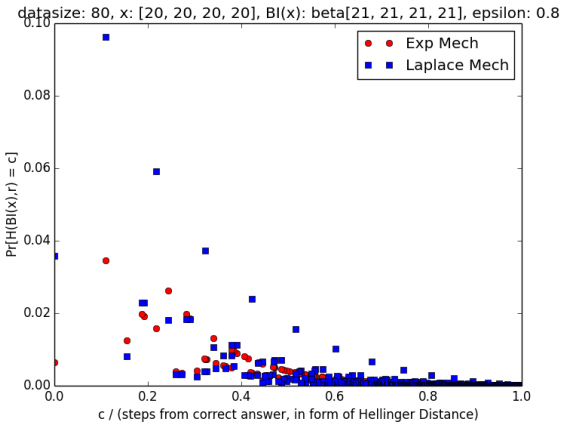



(b) discrete probabilities wrt. hellinger distance

Figure 6: Settings: Dirichlet prior distribution $\mathsf{DL}(1,1,1)$, observed data $x = (20,20,20)$, $\epsilon = 0.8$ and $\delta = 0.000005$

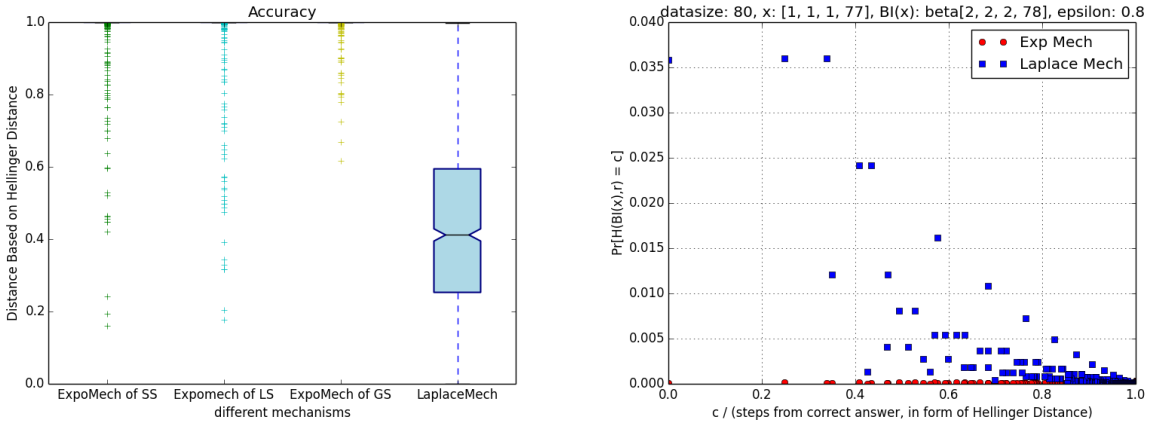



(a) accuracy measurement based on Hellinger distance




(b) discrete probabilities wrt. hellinger distance

Figure 7: Settings: Dirichlet prior distribution $\mathsf{DL}(1,1,1,1)$, observed data $x = (20,20,20,20)$, $\epsilon = 0.8$ and $\delta = 0.000005$




(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 $x = (1,1,1,77)$, $\epsilon = 0.8$ and $\delta = 0.000005$

Here we studied the case when data size $n = 80$ and observed data set $x = (20,20,20,20)$ and plotted these discrete probabilities, shown in Fig. 7. In this plot, SubFig. 7(a) shows that the Laplace Mech are better than Exp Mech very obviously. SubFig (b) shows the Laplace can output most of the good answers with higher probabilities than our exponential mechanism. The advantage in the SubFig. 7(a) is obvious but still not significant in SubFig. 7(b). So, we change the variance of the observed data set to have the next case.

In Fig. 8, we considered an edge case where the variance of the observed data set is small: $x = (1,1,1,77)$ and plotted these discrete probabilities. This plot can show clearly that the Laplace mechanism is much better than our exponential mechanism, both in (a) and (b). In SubFig. 8(b), the average accuracy of Laplace mechanism is conspicuously better than the other three exponential mechanisms. Also, in SubFig. 8(b), the Laplace mechanism can output good answers with much higher probabilities, even though their performances on bad answers are not so obviously. Moreover, no matter good answers or bad answers, their outputting probabilities are very close in our exponential mechanism.

From the theory and experimental analysis above, we obtained some conclusions

1. when the data size and the dimensions of the prior distribution are small (candidate set size $n^{m-1}$ around 5000, our exponential mechanism with smooth sensitivity can do better than Laplace mechanism.

2. when the observed data set is close to center (i.e., the variance is large), the performance of the two mechanisms are close to each other.

3. The Laplace mechanism will beat us when the data size and the dimensions of the prior distribution increasing and when the variance of the observed data set is small.

# 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.