Accuracy Analysis For Three Mechanisms

ABSTRACT

KEYWORDS

Accuracy analysis

1 INTRODUCTION

2 ALGORITHM SETTING UP

Given a prior distribution $\boldsymbol{\beta}_{\text{prior}} = \text{beta}(\alpha, \beta)$ and a sequence of n observations $\mathbf{x} \in \{0,1\}^n$, we define the follwing set:

$$\mathcal{R}_{\text{post}} \equiv \{\text{beta}(\alpha', \beta') \mid \alpha' = \alpha + \Delta\alpha, \beta' = \beta + n - \Delta\alpha\}$$

where $\Delta\alpha$ is as defined in Section **??**. Notice that $\mathcal{R}_{\text{post}}$ has $n+1$ elements, and the Bayesian Inference process will produce an element from $\mathcal{R}_{\text{post}}$ that we denote by $\text{BI}(\mathbf{x})$ – we don't explicitely parametrize the result by the prior, which from now on we consider fixed and we denote it by $\boldsymbol{\beta}_{\text{prior}}$.

2.1 Baseline Approaches

2.1.1 Exponential Mechanism. Exponential mechanism $\mathcal{M}_E(x, u, \mathcal{R}_{\text{post}})$ samples a element from the candidate set $\mathcal{R}_{\text{post}} = \{r_1, r_2, \cdots r_n\}$ with probability proportional to $exp(\frac{\epsilon u(x,r)}{2GS})$:

$$\Pr_{z \sim \mathcal{M}_E(x,u,\mathcal{R}_{\text{post}})}[z = r] = \frac{exp(\frac{\epsilon u(x,r)}{2GS})}{\Sigma_{r' \in \mathcal{R}}\ exp(\frac{\epsilon u(x,r')}{2GS})},$$

where $u(x,r)$ is the Hellinger scoring function over candidates, $\mathcal{H}(\text{BI}(\mathbf{x}), r)$, and $GS(x)$ is the global sensitivity calculated by:

$$GS = \max_{\{|\mathbf{x},\mathbf{x}'|\leq 1; \mathbf{x},\mathbf{x}' \in \mathcal{X}^n\}} \max_{\{r \in \mathcal{R}\}} |\mathcal{H}(\text{BI}(\mathbf{x}), r) - \mathcal{H}(\text{BI}(\mathbf{x}'), r)|$$

Exponential mechanism is $\epsilon-$differential privacy[1].

2.1.2 Exponential Mechanism with Local Sensitivity. Exponential mechanism with local sensitivity $\mathcal{M}_E^{local}(x, u, \mathcal{R}_{\text{post}})$ 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 $exp(\frac{\epsilon u(x,r)}{2LS(x)})$:

$$\Pr_{z \sim \mathcal{M}_E^{local}(x,u,\mathcal{R}_{\text{post}})}[z = r] = \frac{exp(\frac{\epsilon u(x,r)}{2LS(x)})}{\Sigma_{r' \in \mathcal{R}}\ exp(\frac{\epsilon u(x,r')}{2LS(x)})},$$

where local sensitivity is calculated by:

$$LS(\mathbf{x}) = \max_{\mathbf{x}' \in \mathcal{X}^n : \mathbf{adj}(\mathbf{x},\mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\text{BI}(\mathbf{x}'), r) - \mathcal{H}(\text{BI}(\mathbf{x}'), r)|.$$

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

2.1.3 Laplace Mechanism.

- Release $\text{beta}(\alpha + \lfloor\widetilde{\Delta\alpha}\rfloor_0^n, \beta + n - \lfloor\widetilde{\Delta\alpha}\rfloor_0^n)$.
- $\widetilde{\Delta\alpha} \sim \mathcal{L}(\Delta\alpha, \frac{\Delta\text{BI}}{\epsilon})$.
- $\Delta\text{BI} \equiv \max_{\mathbf{x},\mathbf{x}' \in \{0,1\}^n, ||\mathbf{x}-\mathbf{x}'||_1 \leq 1} ||\text{BI}(\mathbf{x}) - \text{BI}(\mathbf{x}')||_1$.

2.1.4 Improved Baseline Approach. y using sensitivity 1 in 2 dimensions and 2 in higher dimensions. Indeed: we can see the output of the Bayesian inference as a histogram, and $||\text{BI}(\mathbf{x}) - \text{BI}(\mathbf{x}')||_1 \leq 2$.

2.2 Our approach: smoothed Hellinger distance based exponential mechanism

2.2.1 Setting up.

Definition 2.1. The mechanism $\mathcal{M}_{\mathcal{H}}(x)$ outputs a candidate $r \in \mathcal{R}_{\text{post}}$ with probability

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}}[z = r] = \frac{exp\left(\frac{-\epsilon \cdot \mathcal{H}(\text{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}{\sum_{r \in \mathcal{R}_{\text{post}}} exp\left(\frac{-\epsilon \cdot \mathcal{H}(\text{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}.$$

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

$$S(\mathbf{x}) = \max_{\mathbf{x}' \in \{0,1\}^n} \left\{ LS(\mathbf{x}') \cdot e^{-\gamma \cdot d(\mathbf{x},\mathbf{x}')} \right\}, \tag{1}$$

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

This mechanism is based on the basic exponential mechanism [2], with $\mathcal{R}_{\text{post}}$ as the range and $\mathcal{H}(\cdot,\cdot)$ as the scoring function. The difference is that in this mechanism we don't calibrate the noise w.r.t. to the global sensitivity of the scoring function but w.r.t. to the smooth sensitivity $S(\mathbf{x})$ – defined by Nissim et al. [3]– of $\mathcal{H}(\text{BI}(\mathbf{x}), \cdot)$.

$\gamma = \gamma(\epsilon, \delta)$ is a function of ϵ and δ to be determined later, and where $LS(\mathbf{x}')$ denotes the local sensitivity at $\text{BI}(\mathbf{x}')$, or equivalently at $\mathbf{x}'$, of the scoring function used in our mechanism.

This mechanism also extends to the Dirichlet-Multinomial system $\text{DL}(\boldsymbol{\alpha})$ by rewriting the Hellinger distance as:

$$\mathcal{H}(\text{DL}(\boldsymbol{\alpha}_1), \text{DL}(\boldsymbol{\alpha}_2)) = \sqrt{1 - \frac{\text{B}(\frac{\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2}{2})}{\sqrt{\text{B}(\boldsymbol{\alpha}_1)\text{B}(\boldsymbol{\alpha}_2)}}},$$

and by replacing the $\mathcal{R}_{\text{post}}$ with set of posterior Dirichlet distributions candidates. Also, the smooth sensitivity $S(\mathbf{x})$ in (1) will be computed by letting $\mathbf{x}'$ range over all the elements in $\mathcal{X}^n$ adjacent to $\mathbf{x}$. Notice that $\mathcal{R}_{\text{post}}$ has $\binom{n+1}{m-1}$ elements in this case. We will denote by $\mathcal{M}_{\mathcal{H}}^D$ the mechanism for the Dirichlet-Multinomial system.

By setting the γ as $\ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$, $\mathcal{M}_{\mathcal{H}}$ is $(\epsilon, \delta)-$differentially private.

3 ACCURACY ANALYSIS

3.1 Laplace Mechanism

Fixing a data set x, we will sample noise $Lap_i = floor(Y)$ where $Y \sim Lap(\frac{2}{\epsilon})$. Since already had the accuracy bound based on the l_1 norm in Laplace mechanism:

$$Pr[|Y| \geq t] = e^{-\frac{t\epsilon}{2}}.$$

So we have the probability $Pr[|Lap_i|] = Pr[|Lap_i| \leq |Y| < |Lap_i| + 1] = e^{-\frac{|Lap_i|\epsilon}{2}} - e^{-\frac{(|Lap_i|+1)\epsilon}{2}}$.

When we are 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. Then the probability of each group of noises are calculated as $Pr[\{|Lap_1|, |Lap_2|, \cdots, |Lap_{m-1}|\}] = Pr[|Lap_1| \leq |Y| < |Lap_1| + 1] \times Pr[|Lap_2| \leq |Y| < |Lap_2| + 1] \times \cdots \times Pr[|Lap_{m-1}| \leq |Y| < |Lap_{m-1}| + 1] = (e^{-\frac{|Lap_1|\epsilon}{2}} - e^{-\frac{(|Lap_1|+1)\epsilon}{2}}) \times (e^{-\frac{|Lap_2|\epsilon}{2}} - e^{-\frac{(|Lap_2|+1)\epsilon}{2}}) \times \cdots \times (e^{-\frac{|Lap_{m-1}|\epsilon}{2}} - e^{-\frac{(|Lap_{m-1}|+1)\epsilon}{2}})$.

3.2 Accuracy Bound for Exponential Mechanism

The accuracy bound of exponential mechanism is provided in [1] as:

$$Pr[u(\mathcal{M}_E(x, u, \mathcal{R}_{\text{post}})) \leq c] \leq \frac{|\mathcal{R}| exp(\frac{\epsilon c}{2GS})}{|\mathcal{R}_{OPT}| exp(\frac{\epsilon OPT_{u(x)}}{2GS})},$$

where $|R|$ is the size of the candidate set, OPT is the optimal candidates, $|R_{OPT}|$ is the number of optimal results.

3.3 Exponential Mechanism with Smooth Sensitivity

We explored three accuracy bounds for our exponential mechanism with smooth sensitivity.

First is the tight bound with very accurate calculation.

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(x)}[\mathsf{H}(\mathsf{BI}(x), z) \geq c] = \sum_{\{z | \mathsf{H}(\mathsf{BI}(x), z) \geq c\}} \frac{e^{\frac{-\epsilon \mathsf{H}(\mathsf{BI}(x), z)}{S(x)}}}{NL_x}.$$

In order to be more efficient, we designed the second accuracy bound which is slightly looser than the first one:

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(x)}[\mathsf{H}(\mathsf{BI}(x), z) \geq c] \leq \frac{|R| \exp(\frac{-\epsilon c}{S(x)})}{NL_x}.$$

In the second bound, we still need to calculate the normaliser every time. So we want make further improvements on efficiency like follows:

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(x)}[\mathsf{H}(\mathsf{BI}(x), z) \geq c] \leq \frac{|R| \exp(\frac{-\epsilon c}{S(x)})}{N(n)},$$

where we replace the NL_x with a value only related to the size of the data. However, we haven't figured out the formula of this $N(n)$.

Moreover, based on the accuracy bound in Sec. 3.2, we can derive a loose bound:

$$Pr[u(\mathcal{M}_{\mathcal{H}}(x)) \leq c] \leq |\mathcal{R}_{\text{post}}| exp(\frac{\epsilon c}{2S(x)}),$$

which has been used in the dilation property proof.

3.4 Accuracy Analysis Comparing with Laplace mechanism

Based on accuracy study above, we are going to have a further study on the accuracy relationship between Laplace mechanism and our exponential mechanism in four aspects: data size, dimensions, prior distribution and data variance.

3.4.1 Accuracy Analysis wrt. Data Size and Distribution Dimensions. In this part, we will analyze the influence of data size 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. For example, in the case where the correct posterior distribution is beta(5, 5), the candidate beta(4, 6) and beta(6, 4) are of 1 steps from beta(5, 5); when correct posterior distribution is DL(5, 5, 5), the candidate DL(4, 5, 3), DL(4, 3, 5), DL(5, 4, 3), DL(5, 3, 4), DL(3, 5, 4) and DL(5, 5, 5) are all of 1 step from DL(5, 5, 5). Under the Hellinger 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:

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

  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 $Lap_i \sim Floor(Lap(\frac{2}{\epsilon}))$ are i.i.d. So, the probabilities of outputting a single candidate in Laplace mechanism can be obtained by:

$$\begin{aligned} Pr[(Lap_1, Lap_2, \cdots, Lap_{m-1})] = Pr[Lap_1 \leq Lap(\frac{2}{\epsilon}) < Lap_1 + 1] \\ \times Pr[Lap_2 \leq Lap(\frac{2}{\epsilon}) < Lap_2 + 1] \times \cdots \\ \times Pr[Lap_{m-1} \leq Lap(\frac{2}{\epsilon}) < Lap_{m-1} + 1], \end{aligned}$$

  where $Pr[Lap_i \leq Lap(\frac{2}{\epsilon}) < Lap_i + 1] = \frac{1}{2} Pr[|Lap_i| \leq Lap(\frac{2}{\epsilon}) < |Lap_i| + 1]$ when $Lap_i > 0$ and $\frac{1}{2} Pr[|Lap_i| - 1 \leq Lap(\frac{2}{\epsilon}) < |Lap_i|]$ else.

  From analysis above, we can see the probabilities of outputting the good answers will not change a lot as the size of the candidate set increasing. Specifically, we can see that the probabilities of outputting good answers are decreasing with speed upper bounded by , 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 beta$(1, 1)$, the observed data set $x = (1, 1, 0, 0, 1, 1, 0, 0)$, it is easy to compute the posterior distribution: beta$(5, 5)$, the probability from the two mechanisms are shown in Tab. 1

    Here, there are only 5 kinds of steps from correct answer. Our mechanism in the second column is clearly better than Laplace mechanism in the third column. When the candidates are close to correct answer (for example, $0, 1, 2$ steps from correct answer), our mechanism can output them with higher probabilities than Laplace mechanism. On the other hand, when candidates are far away from correct answer (for example, $3, 4$ steps), our mechanism can output them with lower probabilities.
  - keep the prior unchanged, beta$(1, 1)$, the observed data set $x = (20, 20)$ (suppose a black box will produce A, B with a certain distribution, after observing this black box continuously for 40 times, we get 20 times A and 20 times B, we can get the posterior distribution beta$(21, 21)$ shown in Tab. 2

    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 bad answers and good answers have very similar outputting probabilities.

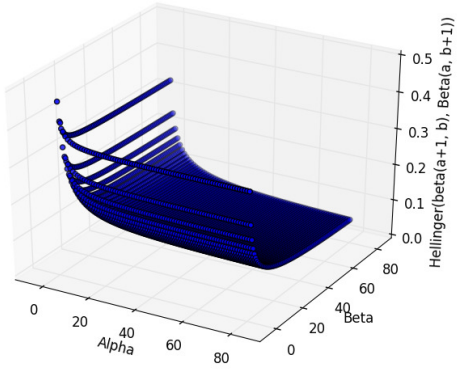

Figure 1: Hellinger distance study

3.4.2 Accuracy Analysis wrt. prior distribution. In this part, we firstly study the local sensitivity of Hellinger distance of different beta distributions as in Fig. 1, in order to have a better understanding of the relationships between accuracy and next two aspects.

As in Fig. 1, the local sensitivity of Hellinger distance will decrease when beta distribution's two parameters get closer (i.e. more uniform) and larger. In the same time, our smooth sensitivity will also decrease based on the Def. 2.1. From accuracy bound in Sec. 3.3, the accuracy will be improved when sensitivity go larger. In consequence, when we increase the prior distribution, the smooth sensitivity in our exponential mechanism will decrease, which means our accuracy will be improved. However, the sensitivity of l_1 norm in Laplace mechanism is fixed regardless the prior distribution. We will study this trade-off in Sec. **??**.

3.4.3 Accuracy Analysis wrt. data variance. Similar as above, when the data variance is small (i.e. the data are uniform), the parameters of posterior distribution will get more uniform (i.e. closer). Based on Fig. 1, the Hellinger distance's sensitivity will get smaller when parameter get closer. As a result, the accuracy will be improved in the same time. We will study this trade-off in Sec. **??**

REFERENCES

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

[2] Frank McSherry and Kunal Talwar. 2007. Mechanism Design via Differential Privacy, In Annual IEEE Symposium on Foundations of Computer Science (FOCS). https://www.microsoft.com/en-us/research/publication/mechanism-design-via-differential-privacy/

[3] 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.

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

Table 1: Probability with Prior beta$(1, 1)$ **and data size** 8

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

Table 2: Probability with Prior beta$(1, 1)$ **and data size** 40