Notes of DP - Bayesian Inference

1 Bayesian Inference Based on Beta-Bernoulli Distribution

The Bayesian inference process is denoted as $\mathsf{BI}(x, prior)$ taking an observed data set $x \in \mathcal{X}^n$ and a prior distribution as input, outputting a posterior distribution $posterior$. For conciseness, when prior is given, we use $\mathsf{BI}(x)$.

2 Algorithm Setting up

For now, we already have a prior distribution $prior$, an observed data set x.

2.1 Exponential Mechanism with Global Sensitivity

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

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

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

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

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

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

where global sensitivity is calculated by:

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

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 $r \in R$ with probability proportional to $exp(\frac{\epsilon u(x,r)}{2\Delta_l u})$:

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

where local sensitivity is calculated by:

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

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

2.3 Exponential Mechanism with Smooth Sensitivity

2.3.1 Algorithm Setting Up

The candidate set and utility function are still the same as before, differ only in the sensitivity. It will output a candidate $r \in R$ with probability proportional to $\exp(\frac{\epsilon u(x,r)}{2S(x)})$:

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

where the sensitivity in mechanism is smooth sensitivity $S(x)$, calculated by:

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

where $\beta = \beta(\epsilon, \delta)$. 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. *for any exponential mechanism $\mathcal{M}_E(x,u,\mathcal{R})$, $\lambda = f(\epsilon,\delta)$, ϵ and $|\delta| < 1$, the sliding property holds:*

$$\underset{z \sim \mathcal{M}_E(x,u,\mathcal{R})}{Pr}[u(r,x) = \hat{s}] \leqslant e^{\frac{\epsilon}{2}} \underset{z \sim \mathcal{M}_E(x,u,\mathcal{R})}{Pr}[u(r,x) = (\Delta + \hat{s})] + \frac{\delta}{2},$$

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

$$\begin{aligned} LHS = \underset{z \sim \mathcal{M}_E(x,u,\mathcal{R})}{Pr}[u(r,x) = \hat{s}] &= \frac{exp(\frac{\epsilon \hat{s}}{2S(x)})}{NL_x} \\ &= \frac{exp(\frac{\epsilon(\hat{s}+\Delta-\Delta)}{2S(x)})}{NL_x} \\ &= \frac{exp(\frac{\epsilon(\hat{s}+\Delta)}{2S(x)} + \frac{-\epsilon\Delta}{2S(x)})}{NL_x} \\ &= \frac{exp(\frac{\epsilon(\hat{s}+\Delta)}{2S(x)})}{NL_x} \cdot e^{\frac{-\epsilon\Delta}{2S(x)})}. \end{aligned}$$

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

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

□

2.3.3 Dilation Property of Exponential Mechanism

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

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

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

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

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

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 $\exp(\frac{\epsilon}{2} \cdot \frac{u(\mathcal{M}_E(x,u,\mathcal{R}))(1-e^{\lambda})}{S(x)})$. Consider the event $G = \{\mathcal{M}_E(x, u, \mathcal{R}) : u(\mathcal{M}_E(x, u, \mathcal{R})) \leqslant \frac{S(x)}{(1-e^{\lambda})}\}$. 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 $r \sim \mathcal{M}_E(x, u, \mathcal{R})$, we can substitute $\hat{s}$ with $u(\mathcal{M}_E(x, u, \mathcal{R}))$. Then, what we need to proof under the case $\lambda > 0$ is:

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

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

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

  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:

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

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

□

3 Experimental Evaluations

3.1 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. 3.1, 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 $r = \mathsf{BI}(x)$.

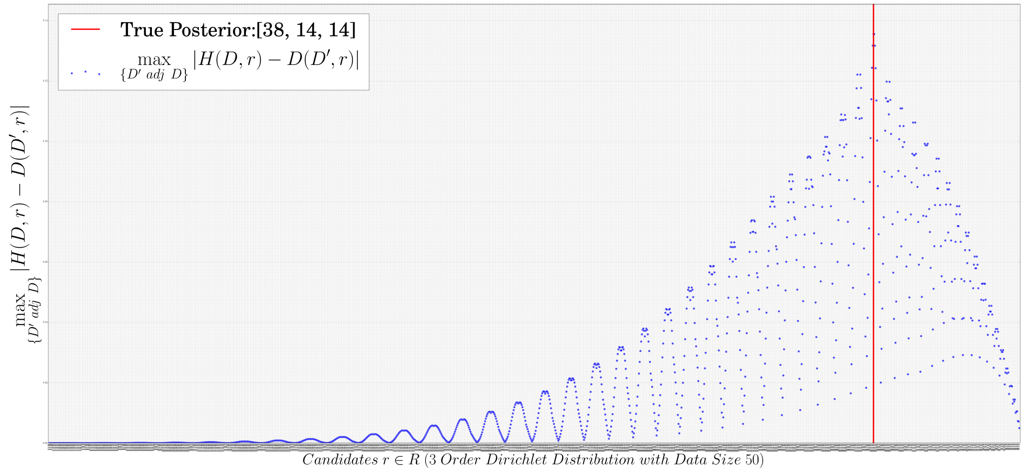

Figure 1: An Experimental Result for Finding the Local Sensitivity efficiently

3.2 Accuracy Study of those algorithm

References

[1] Cynthia Dwork, Aaron Roth, et al. “The algorithmic foundations of differential privacy”. In: *Foundations and Trends® in Theoretical Computer Science* 9.3–4 (2014), pp. 211–407.