CSC2412: Private Multiplicative Weights

Sasho Nikolov
Query Release
Recall the query release problem:

- Workload $Q = \{q_1, \ldots, q_k\}$ of $k$ counting queries

$$Q(X) = \begin{pmatrix} q_1(X) \\ \vdots \\ q_k(X) \end{pmatrix} \in [0,1]^k.$$

- Compute, with $(\varepsilon, \delta)$-DP, some $Y \in \mathbb{R}^k$ so that

$$\max_{i=1}^k |Y_i - q_i(X)| \leq \alpha,$$

with probability $\geq 1 - \beta$. 

In particular,

$$q_i : \mathcal{X} \rightarrow \{0,1\}$$

$$q_i(X) = \frac{1}{n} \sum_{i=1}^n q_i(x)$$

where $X = \{x_1, \ldots, x_n\}$. 
Motivating example

\(\ell\)-wise marginals queries:

- \(\mathcal{X} = \{0, 1\}^d\) \(\text{i.e. } d\) binary attributes
- a query \(q_{S,a}\) for any \(S = \{i_1, \ldots, i_\ell\} \subseteq [d]\) and \(a = (a_{i_1}, \ldots, a_{i_\ell})\):

\[
q_{S,a}(x) = \begin{cases} 
1 & x_{ij} = a_{ij} \quad \forall i \in S \\
0 & \text{otherwise}
\end{cases}
\]

E.g., “smoker and female?”, “smoker and over 30?”, “smoker and heart disease?”, etc.

\[Q_\ell = \text{workload of all } \ell\text{-wise marginal queries on } \{0, 1\}^d \]

\[|Q_\ell| = \binom{d}{\ell} \cdot 2^\ell \approx \left(\frac{2d}{\ell}\right)^\ell\]
What do we know?

\[ \varepsilon\text{-DP: } \frac{d! \cdot \log d}{\varepsilon p} \]

For \( \varepsilon \)-wise marg.

\[ n \gg c \]

Using the Laplace noise mechanism, we can answer \( k \) counting queries with noise \( \leq \delta \) with prob \( \geq 1 - \beta \) when 
\[ n \gg \frac{k \log(k/\beta)}{\varepsilon \delta} \]

\[ \text{(\( \varepsilon, \delta \))-DP: } \frac{d^{1/2} \cdot \log d}{\varepsilon p} \cdot \sqrt{\log(k/\beta)} \]

\[ n \gg \frac{\sqrt{k \log(k/\beta)} \cdot \sqrt{\log(k/\beta)}}{\varepsilon p} \]
We will see an algorithm that achieves:

- under $\varepsilon$-DP, error $\alpha$ with probability $1 - \beta$ when
  \[
  n \gg \frac{\log(k) \log(|\mathcal{X}|)}{\alpha^3 \varepsilon}.
  \]

- under $(\varepsilon, \delta)$-DP, error $\alpha$ with probability $1 - \beta$ when
  \[
  n \gg \frac{\log(k) \sqrt{\log(|\mathcal{X}|) \log(1/\delta)}}{\alpha^2 \varepsilon}.
  \]
Learning a distribution
We can think of $X = \{x_1, \ldots, x_n\}$ as a probability distribution $p$:

\[
\mathbb{P}_{x \sim p}(x = y) = \frac{\{i : x_i = y\}}{n}
\]

Then, for any counting query $q : \mathcal{X} \to \{0, 1\}$,

\[
q(X) = \frac{1}{n} \sum_{i=1}^{n} q(x_i) = \sum_{x \in \mathcal{X}} q(x) \cdot \frac{\{i : x_i = x\}}{n} = \mathbb{E}_{x \sim p} q(x) \Rightarrow q(p)
\]

i.e. $q(X)$ = expectation of $q$ under the empirical distribution of $X$.
Learning a distribution

Task: Learn an approximation $\hat{p}$ of the empirical distribution $p$ such that

$$\forall q \in Q : |q(\hat{p}) - q(p)| \leq \alpha.$$
Bounded mistake learner

Distribution learning algorithm $U$:

- takes a $\hat{p}$ and $q$ such that $q(\hat{p}) - q(p) > \alpha$
- returns a new distribution $\hat{p}_0 = U(q, \hat{p})$

Suppose that $\hat{p}_0$ is uniform over $X$ and $\hat{p}_t = U(\hat{p}_{t-1}, q_t)$.

$U$ makes at most $L$ mistakes if any such sequence $\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_l$ must have $l \leq L$.

After making $L$ mistakes (and $L$ improvements), $\hat{p}_l$ must be accurate for all $q$. 
Multiplicative Weights Learner

Theorem

There exists a distribution learner $U$ that makes

$$L \leq \frac{4 \ln |\chi|}{\alpha^2}$$

mistakes.
**The Learner**

**Multiplicative Weight Update Algorithm**

Reminder: \( q(\hat{p}) - q(p) > \lambda \)

I.e. \( \hat{p} \) gives too much weight to \( x \) st. \( q(x) = 1 \)

\[
q(\hat{p}) = \mathbb{E}_{x \sim \hat{p}} q(x)
\]

\( \hat{p}(x) = \text{prob of } x \text{ under } \hat{p} \)

\[
\hat{p}(x) = \text{prob of } x \text{ parameter, to be set later}
\]

**Algorithm**

\[
U(q, \hat{p}):
\]

\[
\forall x \in \mathcal{X} : \hat{p}(x) = \hat{p}(x) e^{-\frac{q(x)}{\lambda}}
\]

\[
\hat{p}'(x) = \frac{\hat{p}(x)}{\sum_{y \in \mathcal{X}} \hat{p}(y)}
\]

return \( \hat{p}' \)

\[
\rightarrow \text{decrease } \hat{p}(y) \text{ if } q(x) = 1
\]

\[
\rightarrow \text{normalize to get a prob distribution}
\]
Why it works

KL-divergence: \( D(p||\hat{p}_t) = \sum_{x \in X} p(x) \log \frac{p(x)}{\hat{p}_t(x)} \)

1. \( D(p||\hat{p}_0) \leq \log |X| \) because \( \hat{p}_0 \) is uniform

\[
D(p||\hat{p}_0) = \sum_{x \in X} p(x) \left( \log (|X|) + \log p(x) \right) = \log |X| - \sum_{x \in X} p(x) \log \frac{1}{|X|} \leq \log |X| - \log |X| = 0
\]

2. \( D(p||\hat{p}_t) \geq 0 \) for all \( t \)

3. \( D(p||\hat{p}_t) - D(p||\hat{p}_{t-1}) \leq \frac{\eta}{2} (q_{t-1}(p) - q_{t-1}(\hat{p}_{t-1})) + \frac{\eta^2}{4} \)

\[
q_{t-1}(\hat{p}_{t-1}) - q_{t-1}(p) > \lambda
\]

Set \( \eta = \lambda \)

\[
-\frac{\eta^2}{2} + \frac{\eta^2}{4} = -\frac{\lambda^2}{4}
\]

\( \eta \leq \sqrt{\frac{\lambda}{2}} \)
Private Multiplicative Weights
Idea for private algorithm

- Start with $t = 0$, $\hat{p}_0$ uniform.

- Private find the most wrongly answered query $q \in Q$
  - If $q(\hat{p}_t) - q(p) < \alpha$, output $\hat{p}_t$ → all queries in $Q$ have error ≤ $\alpha$
  - Else set $\hat{p}_{t+1} = U(\hat{p}_t, q)$ and increase $t$

q is a mistake

terminates after $\leq L = \frac{4 \log 121}{d^2}$ iterations
The algorithm in detail

\[ \hat{p}_0 = \text{uniform over } \mathcal{X} \]

for \( t = 0 \ldots L - 1 \)

- Sample \( q \in Q \) w/ prob \( \propto \exp \left( \frac{n(q(\hat{p}_t) - q(p))}{2\varepsilon_0} \right) \)

\[ Y_t = q(p) + Z_t, \quad Z_t \sim \text{Lap}(0, \frac{1}{\varepsilon_0 n}) \]

if \( q(\hat{p}_t) - Y_t > 2\alpha \)

\[ \hat{p}_{tn} = U(\hat{p}_t - \alpha, q) \]

else Output \( \hat{p}_t \)

\[ \max \text{ error} \leq q(\hat{p}_t) - q(p) + d \]

\[ \leq q(\hat{p}_t) - Y_t + 2d \leq 3d \]

\( \varepsilon_0 \) parameter, to be set in the priv. analysis
Privacy analysis

Approach: bound privacy loss per iteration. Use composition theorem to bound total priv. loss.

Priv. loss per iteration:

\[
\text{Exp mech } \epsilon_0 - \text{DP} \\
\text{Lap mech } \frac{\epsilon_0}{2} - \text{DP}
\]

Total of \( \leq L \) iterations \( \Rightarrow \) total priv. loss \( \leq 2L \epsilon_0 - \text{DP} \)

Set \( \epsilon_0 = \frac{\epsilon}{2L} = \frac{\epsilon L^2}{8 \ln |\mathcal{X}|} \)
Accuracy analysis

1) We want that w/ prob \( \geq 1 - \beta \)

\[ \forall t \quad |Y_t - q(p)| \leq \epsilon \]

5) query in round \( t \)

Laplace mechanism w/ \( \leq L \) adaptive queries

enough to have \( n \geq \frac{\ln(L/\beta)}{\epsilon_0 \epsilon} = \frac{2L \ln(L/\beta)}{3 \epsilon} \approx \frac{L \log(k/\beta)}{3 \epsilon} \)

2) w/ prob \( \geq 1 - \beta \)

at every iteration \( q(P_t) - q(p) \geq \max_{q' \in Q} q'(P_t) - q'(p) - \epsilon \)

\[ n \gg \frac{\log(kL/\beta)}{\epsilon_0 \epsilon} = \frac{2L \log(kL/\beta)}{3 \epsilon} \approx \frac{2 \log(k/\beta)}{3 \epsilon} \]