Differential Privacy Algorithms for Decentralised Multi-Agent RL

N. Hemachandra[1]
Email: nh@iitb.ac.in

Joint work with Prashant Trivedi[2]

[1] Industrial Engineering and Operations Research, IIT Bombay
[2] One Network Enterprises India Private Limited

February 27, 2024

RL Workshop, IISc, 26–28, Feb, 2024

Google Fi suffers data breach, customer info compromised

According to TechCrunch, Google Fi's primary network provider informed the company that suspicious activity had been detected regarding a third-party support system containing a "limited amount" of customer data.

IANS • February 01, 2023, 12:25 IST

Figure: Source: Economic Times

Importance of data privacy

- Protection of personal information, financial records, and health information, etc.
- Threat to organizations such as financial losses, and reputational damage, etc.

Importance of data privacy

- Protection of personal information, financial records, and health information, etc.
- Threat to organizations such as financial losses, and reputational damage, etc.

Challenges in data privacy

- More sophisticated cyberattacks
- Widespread collection and storage of personal information
- Lack of security measures, encryption protocols to safeguard sensitive information

Traditional approaches

- Suppression: removing names, addresses, or any other personal information
- Aggregation: provide summary statistics while obscuring individual-level details
- Perturbation: adding noise or random variation to the data

Traditional approaches

- Suppression: removing names, addresses, or any other personal information
- Aggregation: provide summary statistics while obscuring individual-level details
- Perturbation: adding noise or random variation to the data

Limitations of traditional approaches

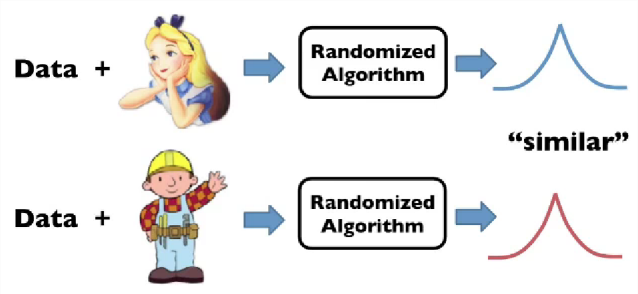
- Re-identification attacks – use of auxiliary information or other datasets
- Privacy and data utility trade-off
- Aggressive anonymization – loss of data utility

Motivation for differential privacy

- Protecting sensitive information in datasets.
- Preventing re-identification of individuals through data analysis.
- Fostering trust between data collectors and individuals.
- Complying with privacy regulations and standards (e.g., Digital Personal Data Protection Act 2023, of India and General Data Protection Regulation (GRDP))

Local differential privacy (LDP)

- Introduced by Dwork et.al. 2006[1]



- Single data point does not change the output

[1] Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." Foundations and Trends ® in Theoretical Computer Science 9.3–4 (2014): 211-407.

Local Differential Privacy

Privacy loss

$$c(o; \mathcal{M}, \textbf{aux}, d, d') := \left| \log \frac{\mathbb{P}(\mathcal{M}(\textbf{aux}, d) = o)}{\mathbb{P}(\mathcal{M}(\textbf{aux}, d') = o)} \right|$$

- Here $\mathcal{M}$ is the randomized mechanism
- **aux** is auxiliary input
- d, d' are neighbouring data points
- o is the outcome

Local Differential Privacy

Privacy loss

$$c(o; \mathcal{M}, \mathbf{aux}, d, d') := \left| \log \frac{\mathbb{P}(\mathcal{M}(\mathbf{aux}, d) = o)}{\mathbb{P}(\mathcal{M}(\mathbf{aux}, d') = o)} \right|$$

- Here $\mathcal{M}$ is the randomized mechanism
- **aux** is auxiliary input
- d, d' are neighbouring data points
- o is the outcome

Local differential privacy (Liao et.al. 2022)

A randomized mechanism $\mathcal{M}$ preserves (ϵ, δ)-LDP if

$$\mathbb{P}(\mathcal{M}(D_u) \in U) \leq e^{\epsilon} \mathbb{P}(\mathcal{M}(D_{u'}) \in U) + \delta, \; U \in \mathcal{U} \tag{1}$$

- $\epsilon \geq 0$, and $\delta \geq 0$ are user given privacy parameters
- $D_u, D_{u'} \in \mathcal{U}$ are the datasets, differing in exactly one component, corresponding to the users u and u'

LDP (outline)

- Consider a task for which mean height, μ, is a crucial input
- Let *D* be a dataset of a cohort
- Height values of *D* need to be protected
- *Anonymise* them.
- One way is to '*add noise*'; say, *U*$[-1, 1]$ (uniform rv, over [-1, 1] interval) to the observed heights

LDP (outline)

- Consider a task for which mean height, μ, is a crucial input
- Let *D* be a dataset of a cohort
- Height values of *D* need to be protected
- *Anonymise* them.
- One way is to '*add noise*'; say, $U[-1, 1]$ (uniform rv, over [-1, 1] interval) to the observed heights

- $\mu = \mu_D + \delta_\mu$
- μ_D is sample mean of heights
- δ_μ is the sample mean of $U[-1, 1]$, is small, *but* not zero
- Consequences?
- Quantify the above error in the estimate?
- May be via concentration inequalities, etc.
- $\mathbb{P}(| \delta_\mu| \leq \epsilon) \geq 1 - \delta$ for ϵ and δ?

Differential privacy for multi-agent system

Multi-agent instance

$$(N, \mathcal{S}, \{\mathcal{A}^i\}_{i\in N}, H, \{r_h^i\}_{i\in N, h\in H}, \{\mathbb{P}_h\}_{h\in H}, \{\mathcal{G}_t\}_{t\geq 0})$$

- State is global information
- Each agent takes independent action
- However, they have a common objective
- Action is a private information; hence, reward is private
- Fixed finite horizon model, total reward criteria

Global state value function

$$V_h^\pi(\mathbf{s}) = \mathbb{E}_\pi\left[\sum_{h'=h}^{H} \bar{r}_{h'}(\mathbf{s}_{h'}, \pi_{h'}(\mathbf{s}_h'))\right]$$

- Here $\bar{r}_{h'}(\mathbf{s}_{h'}, \pi_{h'}(\mathbf{s}_h')) = \frac{1}{n}\sum_{i\in N} r_{h'}^i(\mathbf{s}_{h'}, \pi_{h'}(\mathbf{s}_h'))$

- $\mathcal{G}_t$ is time varying communication network used to exchange the reward parameters $\boldsymbol{w}$ in a decentralized framework
- Particularly, the reward parameters are exchanged via $\mathcal{G}_t$
- Thus, our MARL framework is fully decentralized

Global state-action value function

$$Q_h^{\pi}(\boldsymbol{s},\boldsymbol{a}) = \mathbb{E}_{\pi}\left[\bar{r}_h(\boldsymbol{s},\boldsymbol{a}) + \sum_{h'=h+1}^{H} \bar{r}_{h'}(\boldsymbol{s}_{h'}, \pi_{h'}(\boldsymbol{s}'_h))\right]$$

Multi-agent local differential privacy

A randomized mechanism $\mathcal{M}$ preserves (ϵ, δ) MA-LDP if

$$\mathbb{P}(\mathcal{M}(\mathbf{D}_u) \in U) \leq e^{\epsilon}\mathbb{P}(\mathcal{M}(\mathbf{D}_{u'}) \in U) + \delta, \ U \in \mathcal{U}. \quad (2)$$

- $\epsilon \geq 0$, and $\delta \geq 0$ are user given privacy parameters
- Here $\mathbf{D}_u = (D_u^1, D_u^2, \cdots D_u^n) \in \mathcal{U}$ and $\mathbf{D}_{u'} = (D_{u'}^1, D_{u'}^2, \ldots, D_{u'}^n) \in \mathcal{U}$
- D_u^i and $D_{u'}^i$ differs at exactly one component
- User $u \in K$ is different from agent $i \in N$

Learning objective

Objective 1

Design a decentralized MA-LDP algorithm such that following regret over K episodes is minimized

$$R_K = \sum_{k=1}^{K} \left(\frac{1}{n} \sum_{i \in N} \{ V_1^{\star,i}(\boldsymbol{s}_1^k) - V_1^i(\boldsymbol{s}_1^k) \} \right) \qquad (3)$$

$V_1^{\star,i}(\boldsymbol{s}_1^k)$ is a global value function in the eyes of agent i with full privacy (no privacy loss with full confidence)

We design a decentralized MA-LDP algorithm with sub-linear regret !

Noise adding mechanisms

- MA-LDP algorithm can handle any noise adding mechanisms
- We use Gaussian, Laplace, Uniform, and Bounded Laplace
- Gaussian and Laplace – unbounded supports

Noise adding mechanisms

- MA-LDP algorithm can handle any noise adding mechanisms
- We use Gaussian, Laplace, Uniform, and Bounded Laplace
- Gaussian and Laplace – unbounded supports
- Unbounded support noise mechanisms inject high noise to the sensitive information, though with low probability
- Loss of data utility – motivates the bounded noise mechanisms
- Uniform and bounded Laplace mechanisms
- Bounded support of noise models capture finite precision arithmetic of computers

Learning objective

Objective 2

How does privacy and regret change with the noise distribution support?

- Bounded mechanisms preserve the MA-LDP privacy
- We show that our MA-LDP algorithm has sub-linear regret.
- Regret depends on the end points and the parameters of the noise distribution support!

Function approximations

- To address large state and action spaces

Linearity assumption

$$\mathbb{P}(\boldsymbol{s}'|\boldsymbol{s},\boldsymbol{a}) = \langle \phi(\boldsymbol{s}'|\boldsymbol{s},\boldsymbol{a}), \boldsymbol{\theta}^\star \rangle \text{ for any triplet } (\boldsymbol{s}',\boldsymbol{a},\boldsymbol{s}) \in \mathcal{S} \times \mathcal{A} \times \mathcal{S}$$

Notation

$$\mathbb{P}V(\boldsymbol{s},\boldsymbol{a}) = \sum_{\boldsymbol{s}' \in \mathcal{S}} \langle \phi(\boldsymbol{s}'|\boldsymbol{s},\boldsymbol{a}), \boldsymbol{\theta}^\star \rangle \, V(\boldsymbol{s}') = \langle \phi_V(\boldsymbol{s},\boldsymbol{a}), \boldsymbol{\theta}^\star \rangle, \;\; \forall \;\; \boldsymbol{s},\boldsymbol{a}$$

- Ridge regression to get optimal model parameters $\boldsymbol{\theta}^\star$

Linearity of reward functions

$$\bar{r}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}^\star) = \langle \psi(\boldsymbol{s},\boldsymbol{a}), \boldsymbol{w}^\star \rangle, \;\; \forall \;\; \boldsymbol{s},\boldsymbol{a}$$

- The reward parameterization preserves the privacy of rewards (not the LDP objective!)

Equivalence of optimization problems

- The least square minimizer of the reward function

$$\min_{\boldsymbol{w}} \ \mathbb{E}_{\boldsymbol{s},\boldsymbol{a}}[\bar{r}(\boldsymbol{s},\boldsymbol{a}) - \bar{r}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w})]^2. \qquad \text{(OP 1)}$$

- The above optimization problem is equivalently characterized as

$$\min_{\boldsymbol{w}} \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{s},\boldsymbol{a}}[r^i(\boldsymbol{s},\boldsymbol{a}) - \bar{r}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w})]^2. \qquad \text{(OP 2)}$$

- OP1, and OP2 has same stationary points
- A key aspect of the decentralized algorithm

Reward parameters update

$$\widetilde{\boldsymbol{w}}_t^i \leftarrow \boldsymbol{w}_t^i + \gamma_t \cdot [r_t^i(\cdot,\cdot) - \bar{r}(\cdot,\cdot;\boldsymbol{w}_t^i)] \cdot \nabla_{\boldsymbol{w}} \bar{r}(\cdot,\cdot;\boldsymbol{w}_t^i)$$

$$\boldsymbol{w}_{t+1}^i = \textstyle\sum_{j\in N} l_t(i,j) \widetilde{\boldsymbol{w}}_t^j$$

- $l_t(i,j)$ is the (i,j)-th entry of communication graph/matrix
- Result: $\mathbf{w}_t^i \to \mathbf{w}^\star$ almost surely for every agent $i \in N$

Some comments

- Our MA-LDP is decentralized algorithm [2]:
- Each agent is independently taking the action
- Agents' reward is a private information, and hence not known to other agents
- The reward function is parameterized and the parameters are shared across the agents
- This doesn't effect the reward and action privacy
- The sensitive information is preserved by injecting the noise

[2] Kaitang Zhang et. al. Fully decentralized multi-agent reinforcement learning with networked agents. ICML 2018.

Modified Bellman equation

- Let $V^i(\cdot)$ and $Q^i(\cdot,\cdot)$ be the estimate of global $V(\cdot)$ and $Q(\cdot,\cdot)$ by agent i

Modified Bellman equation

$$Q_h^{\star,i}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i) = \bar{r}_h(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i) + \mathbb{P}_h V_{h+1}^{\star,i}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i);$$

$$V_{h+1}^{\star,i}(\boldsymbol{s};\boldsymbol{w}_{k,h}^i) = \max_{\boldsymbol{a}\in\mathcal{A}} Q_h^{\star,i}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i); \quad V_{H+1}^{\star,i}(\boldsymbol{s};\boldsymbol{w}_{k,h}^i) = 0$$

- $Q_h^{\star,i}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i)$, $\bar{r}_h(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i)$ and $V_h^{\star,i}(\boldsymbol{s};\boldsymbol{w}_{k,h}^i)$ are continuous functions of $\boldsymbol{w}_{k,h}^i$

Result

$$Q_h^{\star,i}(\boldsymbol{s},\boldsymbol{a};\boldsymbol{w}_{k,h}^i) \to Q_h^\star(\boldsymbol{s},\boldsymbol{a}) \text{ and } V_h^{\star,i}(\boldsymbol{s};\boldsymbol{w}_{k,h}^i) \to V_h^\star(\boldsymbol{s}), \text{ for all } i \in N$$

MA-LDP algorithm design

- MA-LDP works in episodes
- Each user/episode receives the information from server
- The server updates the model parameters using the anonymized information

$$\hat{\boldsymbol{\theta}}^i_{k+1,h} \leftarrow (\Sigma^i_{k+1,h})^{-1} u^i_{k+1,h} \tag{4}$$

- Here Σ^i and u^i are anonymized sensitive information
- Server sends model parameters $\hat{\theta}^i$ to next user

MA-LDP algorithm design

- MA-LDP works in episodes
- Each user/episode receives the information from server
- The server updates the model parameters using the anonymized information

$$\hat{\boldsymbol{\theta}}^i_{k+1,h} \leftarrow (\Sigma^i_{k+1,h})^{-1} u^i_{k+1,h} \tag{4}$$

- Here Σ^i and u^i are anonymized sensitive information
- Server sends model parameters $\hat{\theta}^i$ to next user
- User, on the other hand, updates $Q^i_{k,h}$ according to the backward induction algorithm
- Each agent thus take action

$$\boldsymbol{a}^i_{k,h} \leftarrow arg \max_{\boldsymbol{a} \in \mathcal{A}^i} \min_{\boldsymbol{a}^{-i} \in \mathcal{A}^{-i}} Q^i_{k,h}(\boldsymbol{s}_{k,h}, \boldsymbol{a}, \boldsymbol{a}^{-i})$$

- The reward function parameters are shared via communication network to preserve the privacy of rewards

MA-LDP algorithm design

- The anonymized information is send to the server
- This server is different from the centralized server used in centralized MARL
- Server performs the following updates
  - $\Lambda^i_{k+1,h} \leftarrow \Lambda^i_{k,h} + \Delta\Lambda^i_{k,h}$
  - $u^i_{k+1,h} \leftarrow u^i_{k,h} + \Delta u^i_{k,h}$
  - $\Sigma^i_{k+1,h} \leftarrow \Lambda^i_{k+1,h} + \eta I$
  - $\hat{\theta}^i_{k+1,h} \leftarrow (\Sigma^i_{k+1,h})^{-1} u^i_{k+1,h}$
- Here,

$$\Delta\mathbf{\Lambda}^i_{k,h} \leftarrow \phi_{V^i_{k,h+1}}(\boldsymbol{s}_{k,h}, \boldsymbol{a}_{k,h})\phi_{V^i_{k,h+1}}(\boldsymbol{s}_{k,h}, \boldsymbol{a}_{k,h})^\top + \mathbf{W}^i_{k,h}$$

$$\Delta\mathbf{u}^i_{k,h} \leftarrow \phi_{V^i_{k,h+1}}(\boldsymbol{s}_{k,h}, \boldsymbol{a}_{k,h})V^i_{k,h+1}(s_{k,h+1}) + \boldsymbol{\xi}^i_{k,h}$$

Regret and privacy gurantees

- MA-LDP algorithm preserves LDP for various noise mechanisms
- For Gaussian mechanism MA-LDP is (ϵ, δ) private
- For Laplace it is $(\epsilon, 0)$

Regret and privacy gurantees

- MA-LDP algorithm preserves LDP for various noise mechanisms
- For Gaussian mechanism MA-LDP is (ϵ, δ) private
- For Laplace it is $(\epsilon, 0)$

- We introduce uniform and bounded Laplace mechanisms
- These preserve $(0, \delta)$, and $(\epsilon, 0)$ privacy respectively
- Thus, these noise mechanisms cover whole spectrum of the privacy guarantees

Regret and privacy gurantees

- MA-LDP algorithm preserves LDP for various noise mechanisms
- For Gaussian mechanism MA-LDP is (ϵ, δ) private
- For Laplace it is $(\epsilon, 0)$

- We introduce uniform and bounded Laplace mechanisms
- These preserve $(0, \delta)$, and $(\epsilon, 0)$ privacy respectively
- Thus, these noise mechanisms cover whole spectrum of the privacy guarantees

- For each of the noise mechanisms – regret is sub-linear in K
- It is super-linear (not quadratic) in n, i.e., scales well with n
- For bounded Laplace, regret depends on the endpoint of the support and the distribution parameters

Main results

Mechanism	Privacy	Order of Regret
Gaussian	(ϵ, δ)	$\widetilde{\mathcal{O}}((nd)^{5/4}H^{7/4}T^{3/4}\log(ndT/\alpha)(\log(H/\delta))^{1/4}\sqrt{1/\epsilon})$
Laplace	$(\epsilon, 0)$	$\widetilde{\mathcal{O}}((nd)^{5/4}H^{7/4}T^{3/4}\log(ndT/\alpha)\sqrt{1/\epsilon})$
Uniform	$(0, \delta)$	$\widetilde{\mathcal{O}}((nd)^{5/4}H^{7/4}T^{3/4}\log(ndT/\alpha)(\log(H/\delta))^{1/4}$
Bounded Laplace	$(\epsilon, 0)$	$\widetilde{\mathcal{O}}((nd)^{5/4}\zeta^{1/4}H^{1/4}T^{3/4}\log(ndT/\alpha))$

Table: Privacy guarantees and the order of regret for different noise adding mechanisms. ζ denotes the variance of bounded Laplace distribution.

- ζ is function of end points of the support of bounded Laplace distribution B and ϵ.
- For every noise mechanism, the regret is sub-linear in $T = KH$
- However, it scales super-linearly with the number of agents, n

Comparison of regret for different noise mechanisms

Theorem

If privacy parameters ϵ_1 and ϵ_2 are such that $\epsilon_1 > \epsilon_2$. Then, for both the Gaussian and Laplace mechanisms we have that $R_K(\epsilon_1) < R_K(\epsilon_2)$.

Comparison of regret for different noise mechanisms

Theorem

If privacy parameters ϵ_1 and ϵ_2 are such that $\epsilon_1 > \epsilon_2$. Then, for both the Gaussian and Laplace mechanisms we have that $R_K(\epsilon_1) < R_K(\epsilon_2)$.

Theorem

Let $R_K^G(\epsilon), R_K^L(\epsilon)$ be the cumulative regret of the Gaussian and Laplace mechanism respectively with privacy parameters ϵ, δ, and $H > 2$. Then, $R_K^G(\epsilon) > R_K^L(\epsilon)$.

Regret of bounded Laplace

- We construct a BL distribution with parameter b and support $[-B, B]$

$$f_{\mathcal{BL}}(x;b) = \begin{cases} \frac{\exp(-|x|/b)}{2b(1-\exp(-B/b))}, & \forall\, x \in [-B, B] \\ 0, & \text{otherwise.} \end{cases}$$

- The regret is sub-linear in $T = KH$ and super-linear in n
- Regret of BL is either same or on par with the Laplace when $B = O(b^\gamma)$ for $\gamma \in [0, 1]$
- Regret of BL is lower than Laplace if $\gamma > 1$ and $(H^3/\epsilon)^{\gamma/2} < 1$

B	R_K^{BL}
$O(b^\gamma), 0 \leq \gamma \leq 1$	$\widetilde{\mathcal{O}}((nd)^{5/4} H^{7/4} T^{3/4} \log(ndT/\alpha))\sqrt{1/\epsilon}$
$O(b^\gamma), \gamma > 1$	$\widetilde{\mathcal{O}}((nd)^{5/4} H^{7/4} H^{3\gamma/2} T^{3/4} \log(ndT/\alpha))\sqrt{1/\epsilon^{\gamma+1}}$

Table: Regret bound for BL mechanism. MA-LDP algorithm with BL mechanism offers the same order of regret as that of the Laplace mechanism when $B = O(b^\gamma)$ for $\gamma \in [0, 1]$. Terms in red involve γ.

Proof Sketch

- Privacy analysis
  - Show that privacy loss is bounded by ϵ with high probability δ
  - ϵ, δ depends on the noise mechanism used
- Regret analysis
  - Transition probability estimators are within specified range of true optimal parameters (Lemma 1, next slide)
  - $Q^{\star,i}$ is indeed a good optimistic estimator (Lemma 2, next slide)
  - Decomposition of regret and bounding each term
- The regret and privacy comparison across noise adding mechanisms

Proof sketch

Lemma 1 (informal statement)

For all $i \in N$, with probability at least $1 - \alpha/2$, we have

$$||(\Sigma^{i}_{k,h})^{1/2}(\hat{\boldsymbol{\theta}}^{i}_{k,h} - \boldsymbol{\theta}^{\star}_{h})|| \leq \beta_k$$

- Here β_k are identified according to the noise mechanism used
- This proves that the optimistic estimators of the probability function are with a specified range of the true optimal parameters

Proof sketch

Lemma 1 (informal statement)

For all $i \in N$, with probability at least $1 - \alpha/2$, we have
$$||(\Sigma^i_{k,h})^{1/2}(\hat{\boldsymbol{\theta}}^i_{k,h} - \boldsymbol{\theta}^\star_h)|| \leq \beta_k$$

- Here β_k are identified according to the noise mechanism used
- This proves that the optimistic estimators of the probability function are with a specified range of the true optimal parameters

Lemma 2 (informal statement)

For all $i \in N$, we have $Q^{\star,i}_h(\boldsymbol{s}, \boldsymbol{a}) \leq Q^i_{k,h}(\boldsymbol{s}, \boldsymbol{a})$ and $V^{\star,i}_h(\boldsymbol{s}) \leq V^i_{k,h}(\boldsymbol{s})$

- The above lemma shows that the $Q^{\star,i}$ is a good optimistic estimator

Experiments

- The network consists of $\{s_{in}, 1, 2, \ldots, q, g\}$ nodes
- Actions $\mathcal{A}^i = \{-1, 1\}^{d-1}$, $d \geq 2$
- Objective: to reach the goal node while maximizing the overall reward
- Reward of $5/1000$ for any action in s_{in}
- Reward of 1000 for any action in g
- Reward of 0 for any action in any other node

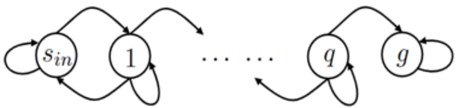


Figure: The MDP problem instance that we consider

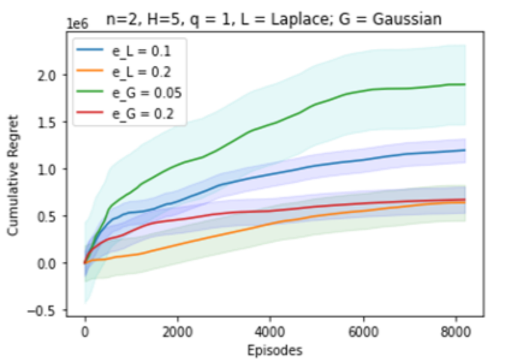


Figure: Cumulative regret with number of episodes for the Laplace and Gaussian mechanism with 5% error bands. Codes are available here.

Discussions

- An observation: If the support of bounded noise distribution is picked appropriately, the regret is lower than the unbounded support noise mechanism
- Injecting a bounded noise is often sufficient for LDP without substantially affecting the nature of the regret
- Bounded noise captures the realistic finite machine precision

Discussions

- An observation: If the support of bounded noise distribution is picked appropriately, the regret is lower than the unbounded support noise mechanism
- Injecting a bounded noise is often sufficient for LDP without substantially affecting the nature of the regret
- Bounded noise captures the realistic finite machine precision
- Another observation: Our regret bound is just (not quadratic) super-linear in the number of agents and feature dimensions
- Scope for using better optimistic estimators of the state-action value functions to improve the bounds
- Studying the bounded support noise mechanism with lower regret bounds with low noise values would be interesting

Thank You!