Generalized Linear Bandits with Local Differential Privacy

Zhipeng Liang

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Yuxuan Han[1]*, Zhipeng Liang[2]*, Yang Wang[1,2], and Jiheng Zhang[1,2]
Department of Mathematics[1]
Department of Industrial Engineering and Decision Analytics[2]
The Hong Kong University of Science and Technology

GitHub: liangzp/LDP-Bandit

arXiv: 2106.03365v1

Online Learning

A general framework of Multi-arm contextual bandit:

1. X_t: individual-specific context.
2. θ_a: parameter of action $a \in [1, \ldots, K]$.
3. Objective: maximize reward $r_t = \mu(X_t, \theta_{a_t}) + \varepsilon_t$.

$$\text{Reg}(T) = \sum_{t=1}^{T} \Big[\mu(X_t, \theta_{a_t^*}) - \mu(X_t, \theta_{a_t})\Big].$$

Key: **exploration v.s. exploitation**

New challenge: How to keep X_t (and even a_t and r_t) private?

Bandits Algorithm

$Y_t = X_t^T \theta_i + \epsilon_{\{i,t\}}$

X_t for users features and θ_i for item i's feature greedy.pdf (stanford.edu), web.stanford.edu/~bayati/papers/lassoBandit.pdf
At each time a X_t represent a users arrive and is offered a set of candidate items $\{\theta_1, \dots, \theta_K\}$

$Y_t = X_{t,i}^T \theta_{\mathrm{j}}$

$X_{t,i}$ for items features and θ_j for user j's feature [1401.8257] Online Clustering of Bandits (arxiv.org).
At each time a set of contexts $\{x_{t,1}, \dots, x_{t,K}\}$ represent a set of candidate items is offered and a user index by j arrive

$Y_t = X_{t,i}^T \theta$

$X_{t,i}$ simultaneously involve the modeling of users and items, 2004.06321.pdf (arxiv.org),

Rigorous Explanations

As there are different (but equivalent) formulations of contextual bandits, we briefly discuss the meaning of the above abstract quantities and how they arise in practice. In general, at each round t,

an individual characterized by v_t (a list of characteristics associated with that individual) becomes available. When the decision maker decides to apply action a_t to this individual, a reward $y_t(v_t, a_t)$, which depends (stochastically) on both v_t and a_t, is obtained. In practice, for both modelling and computational reasons, one often first featurizes the individual characteristics and the actions. In particular, with sufficient generality, one assumes $\mathbf{E}[y_t(v_t, a_t) \mid v_t, a_t] = g_\theta(\phi(v_t, a_t))$, where $g_\theta(\cdot)$ is the parametrized mean reward function and $\phi(v_t, a_t)$ extracts the features from the given raw individual characteristics v_t and action a_t. In the above formulation, as is standard in the literature, we assume the feature map $\phi(\cdot)$ is known and given and $x_{t,a} = \phi(v_t, a)$. Consequently, we directly assume access to contexts $\{x_{t,a} \mid a \in [K]\}$. Note that the linear contextual bandits setting then corresponds to $g_\theta(\cdot)$ is linear.

Local Differential Privacy

- Local differential privacy (LDP), a strigent notion of privacy with provable privacy guarrantee.
- Academic research Rubinstein et al. (2009); Dwork and Lei (2009); Wasserman and Zhou (2010); Smith (2011); Chaudhuri et al. (2011).
- Industry adoption Erlingsson et al. (2014); Ding et al. (2017); Tang et al. (2017).

Local Differential Privacy

Definition (ε-LDP Dwork et al. (2006); Dwork and Roth (2013))

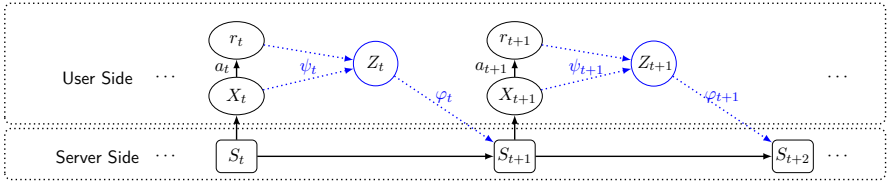
A mechanism $M : \mathcal{X} \to \mathcal{Z}$ is (ε, δ)-LDP if for every subset $C \subset \mathcal{Z}$ and any two different $x, x' \in \mathcal{X}$, we have

$$\mathbb{P}(M(x) \in C) \leq e^{\varepsilon}\mathbb{P}(M(x') \in C) + \delta.$$

LDP

Every user involved in this algorithm is guaranteed that anyone else can only access her context (and related information) with limited advantage over a random guess.

Server-User Communication Protocol



User side:

1. Receive $S_t = (\psi_{t-1}(X_{t-1}, a_{t-1}, r_{t-1}), \ldots, \psi_1(X_1, a_1, r_1))$ from server.
2. Make decision a_t based on (X_t, S_t) and receive corresponding reward r_t.
3. Generate $Z_t = \psi_t(X_t, a_t, r_t)$ and send it to the server.

Server side:

1. Receive Z_t and merge it into S_t to generate S_{t+1}.
2. Send S_{t+1} to the $t+1$-th user.

Results (Non-privacy Bandits)

- In single param settings, existing non-privacy **worst-case** bounds are $\tilde{O}(\sqrt{T})$ Kannan et al. (2018); Sivakumar et al. (2020); Han et al. (2020).
- In single param settings, existing non-privacy bounds under **margin-condition** are $\tilde{O}(\log(T))$ Han et al. (2020).
- In multi param setting, the **margin-dependent** bound is considered in Bastani and Bayati (2020) with regret bound $O(\log(T))$.

Results (LDP Mechanism)

- Dedicated LDP mechanisms have been designed for privacy guarantee in offline learning (unsupervised and supervised)

  Key idea: Inject randomness into the input without distructing data analysis.
- LDP requires the privacy protection mechanism to fluctuate only sightly (in probability) with any change in its input
  - to prevent attackers from recovering the private information of an individual
  - contradicts the goal of personalized decision making
- Shariff and Sheffet (2018) shows that contextual bandit under differential privacy has a lower bound for regret $\Omega(T)$.

Challenges

To achieve sublinear regret: JDP (joint differential privacy).

- still prevents the privacy leakage of the t-th user from the information released after time period t
- no longer requires any privacy constraint on the t-th output with respect to the t-th user

Bandits Learning with JDP/LDP guarantee:

- Shariff and Sheffet (2018) design a variant of UCB with JDP guarantee with regret $O(\sqrt{T})$.
- Zheng et al. (2020) combine UCB algorithm with Gaussian Mechanism but can only achieve $O(T^{\frac{3}{4}})$ regret bound with LDP guarantee.

Results

Result	Regret	Context	Parameter	β-Margin
Zheng et al. (2020)	$\tilde{O}(T^{3/4}/\varepsilon)$	Adversary	Both	No Margin
Theorem 3.1	$\tilde{O}(T^{1/2}/\varepsilon)$	Stochastic	Single	No Margin
Theorem 3.3	$O(\log T/\varepsilon^2)$	Stochastic	Single	$\beta = 1$
Theorem 3.3	$\tilde{O}(T^{\frac{1-\beta}{2}}/\varepsilon^{1+\beta})$	Stochastic	Single	$0 \leq \beta < 1$
Theorem 4.4	$O((\log T/\varepsilon)^2)$	Stochastic	Multiple	$\beta = 1$
Theorem 4.4	$\tilde{O}(T^{\frac{1-\beta}{2}}/\varepsilon^{1+\beta})$	Stochastic	Multiple	$0 < \beta < 1$

TABLE: Summary of our main results in (ε, δ)-LDP, where $\tilde{O}(\cdot)$ omits poly-logarithmic factors.

Algorithm 3: Generalized Linear Bandits with LDP

Input: privacy parameters ε, δ, failure probability α

Initialize: $\tilde{V}_0 = 0_{d\times d}, \tilde{u}_0 = 0_d, \tilde{\theta}_0 = \hat{\theta}_1 = 0_d, \zeta = \Theta(1/\sqrt{T}), \sigma = 6\sqrt{2\ln(3.75/\delta)}/\epsilon$

Notations: $\Upsilon_t = \sigma\sqrt{t}(4\sqrt{d} + 2\ln(2T/\alpha)), c_t = 2\Upsilon_t, \beta_t^2 = \tilde{\mathcal{O}}(\frac{C\sigma}{\mu}\sqrt{dt})$

for $t = 1, 2, \ldots$ **do**

  For the local user t**:**

  Receive information $\tilde{V}_{t-1}, \tilde{\theta}_{t-1}, \hat{\theta}_t$ from the server

  Play action $x_t = \text{argmax}_{x\in\mathcal{D}_t} \left\langle \tilde{\theta}_{t-1}, x \right\rangle + \beta_{t-1} \|x\|_{\tilde{V}_{t-1}^{-1}}$

  Observe reward $y_t = g(x_t^\top \theta^*) + \eta_t$, set $z_t = x_t^\top \hat{\theta}_t$.

  Send $x_t x_t^\top + B_t, z_t x_t + \xi_t, \nabla\ell_t(\hat{\theta}_t) + r_t$ to the server, where

  $\ell_t(\theta) = \ell(x_t^\top\theta, y_t), B_t(i,j) \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2), \forall i \leqslant j$, and
  $B(j,i) = B(i,j), \xi_t \sim \mathcal{N}(0_d, \sigma^2 \mathrm{I}_{d\times d}), r_t \sim \mathcal{N}(0_d, C^2\sigma^2\mathrm{I}_{d\times d})$

  For the server:

  Update $\bar{V}_t = \bar{V}_{t-1} + x_t x_t^\top + B_t, \tilde{u}_t = \tilde{u}_{t-1} + z_t x_t + \xi_t$ $\tilde{\theta}_t = \tilde{V}_t^{-1}\tilde{u}_t$, where

  $\tilde{V}_t = \bar{V}_t + c_t \mathrm{I}_{d\times d}$ $\hat{\theta}_{t+1} = \Pi_{\mathcal{W}}\left(\hat{\theta}_t - \zeta(\nabla\ell_t(\hat{\theta}_t) + r_t)\right)$

Algorithm Framework for single-parameter

Single-param:

Each arm share a same parameter

- flexible plug-in

Algorithm 1: LDP Single-parameter Contextual Bandit

Input: Time horizon T; Privacy Level ε, δ.

Initialization: Setting $\hat{\theta}_0 = \mathbf{0}$.

for $t \leftarrow 1$ *to* T **do**

  User side:

    Receive $\hat{\theta}_{t-1}$ from the server.

    Pull arm $a_t = \text{argmax}_{a \in [K]} x_{t,a}^T \hat{\theta}_{t-1}$ and receive r_t.

    Generate Z_t by

$$Z_t = \psi_t(x_{t,a_t}, r_t; \hat{\theta}_{t-1}).$$

  Server side:

    Receive Z_t from the user.

    Update the estimation via

$$\hat{\theta}_t = \varphi_t(Z_1, \ldots, Z_t; \hat{\theta}_{t-1}).$$

end

Privacy Mechanism and Estimator

Ordinary Least Square

$$\psi_t^{OLS}(x_{t,a_t}, r_t; \hat{\theta}_{t-1}) = (M_t, u_t), \tag{1}$$

$$\varphi_t^{OLS}(Z_1, \ldots, Z_t; \hat{\theta}_{t-1}) = \Big(\sum_{i=1}^{t} M_i + \tilde{c}\sqrt{t}I\Big)^{-1} \sum_{i=1}^{t} u_i. \tag{2}$$

Stochastic Gradient Descent

$$\psi_t^{SGD}(x_{t,a_t}, r_t; \hat{\theta}_{t-1}) = \Psi_{\varepsilon,R}\left((\mu(x_{t,a_t}^T \hat{\theta}_{t-1}) - r_t) x_{t,a_t}\right), \tag{3}$$

$$\varphi_t^{SGD}(Z_1, \ldots, Z_t; \hat{\theta}_{t-1}) = \hat{\theta}_{t-1} - \eta_t \psi_t^{SGD}. \tag{4}$$

Main Results – single-parameter

Theorem

With the choice of $\tilde{c} = 2\sigma_{\varepsilon,\delta}(4\sqrt{d} + 2\log(2T/\alpha))$ *in* (2)*, Algorithm (1) with OLS mechanism* ψ_t^{OLS} *and estimator* φ_t^{OLS} *achieve the following regret with probability at least* $1-\alpha$ *for some constant* C*,*

$$Reg(T) \leq C\sqrt{T}(\sigma_{\varepsilon,\delta} d \frac{\sqrt{(d+\log(T/\alpha))\log(KT/\alpha)}}{\kappa_l p_*} + o(1)).$$

Theorem

With SGD mechanism ψ_t^{SGD} *and estimator* φ_t^{SGD} *achieves the following regret with probability at least* $1-\alpha$ *for some constant* C*,*

$$Reg(T) \leq C\sqrt{T}(\frac{d}{\zeta\kappa_l p_* \varepsilon}\log\log(T/\alpha) + o(1)).$$

Main Results – single-parameter

THEOREM

For $\theta \in \mathbb{R}^d$ and an algorithm π, we denote $\mathbb{E}[Reg_\pi(T;\theta)]$ the expectation regret of π when the underlying parameter is θ. When $K = 2$ and $x_{t,a} \sim \mathcal{N}(0, I_d/d)$ are independent over $a \in [K]$, we have for any possible ε-LDP bandit algorithm π,

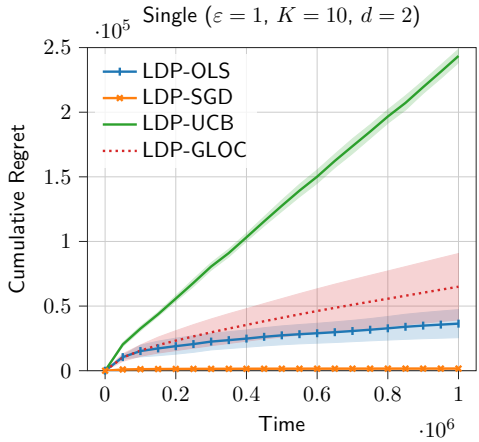
$$\sup_{\theta^\star : \|\theta^\star\|_2 \leq 1} \mathbb{E}[Reg_\pi(T;\theta^\star)] = \Omega(\sqrt{Td}/\varepsilon).$$

Discussion

Our theoretical results restablish the regret bound $O(\sqrt{T}/\varepsilon)$.

- Given the best known $O(T^{3/4}/\varepsilon)$ regret bound of adversarial contextual LDP bandits in Zheng et al. (2020), our $O(\sqrt{T}/\varepsilon)$ result points out a possible gap between stochastic contextual bandits and adversarial contextual bandits under the LDP constraints.
- Given the $O(\sqrt{T} + \frac{\log^{10}(1/\delta)}{\varepsilon^2})$ bound in JDP stochastic contextual dynamic pricing Chen et al. (2020), our $\Omega(\sqrt{T}/\varepsilon)$ result points out a gap between joint differential privacy and local differential privacy in stochastic bandit setting: There is no way to gain local privacy for free, i.e. the privacy factor must influence the total regret in the way of multiplying instead of adding.

Experiment – single-parameter



Algorithm Framework

Multi-param:

Each arm have its own parameter

- warm up stage
- synthetic update
- elimination
- plug-in

Algorithm 2: LDP Multi-parameter Contextual Bandit

Input: Time horizon T; Warm up period length s_0; Privacy Level ε, δ.

Initialization: Setting $\hat{\theta}_{0,i} = 0, i \in [K]$.

for $t \leftarrow 1$ *to* Ks_0 **do**

  User side:

    Receiving $\hat{\theta}_{t-1,1:K}$ from the server.

    Pulling arm $a_t := (t \bmod K) + 1$ and receive r_t.

    Generate and update $Z_{t,i} = \mathbf{1}\{a_t = i\}\psi_t(X_t, r_t; \hat{\theta}_{t-1,i}), i \in [K]$ to the server.

  Server side:

    Receive the update $Z_{t,1:K}$ from the user.

    Re-estimate parameters via $\hat{\theta}_{t,i} := \varphi_t(Z_{1,i}, \ldots, Z_{t,i}), \forall i \in [K]$.

end

for $t \leftarrow Ks_0 + 1$ *to* T **do**

  User side:

    Receive $\hat{\theta}_{t-1,1:K}$ from the server.

    Determine a subset $\hat{K}_t$ of $[K]$ by setting

$$\hat{K}_t := \{a \in [K] : X_t^T \hat{\theta}_{Ks_0,a} > \max_{a \in [K]} X_t^T \hat{\theta}_{Ks_0,a} - \frac{h}{2}\}$$

    Pulling arm $a_t := \operatorname{argmax}_{a \in \hat{K}_t} \mu(X_t^T \hat{\theta}_{t-1,a})$ and receive r_t.

    Generating information for all arms $\{Z_{i,t}\}_{i \in [K]}$ by setting

$$Z_{i,t} = \begin{cases} \psi_t(X_t, r_t; \hat{\theta}_{t-1,i}) & \text{if } a_t = i, \\ \psi_t(0, 0; \hat{\theta}_{t-1,i}) & \text{otherwise.} \end{cases}$$

  Server side:

    Receive the update $\{Z_{i,t}\}_{i \in [K]}$ from the user.

    Re-estimate parameters via

$$\hat{\theta}_{t,i} := \varphi_t(Z_{1,i}, \ldots, Z_{t,i}).$$

end

Main Results – multi-parameter

Theorem

Algorithm 2 with OLS mechanism ψ_t^{OLS} and estimator φ_t^{OLS} achieve the following regret with probability at least $1-\alpha$:

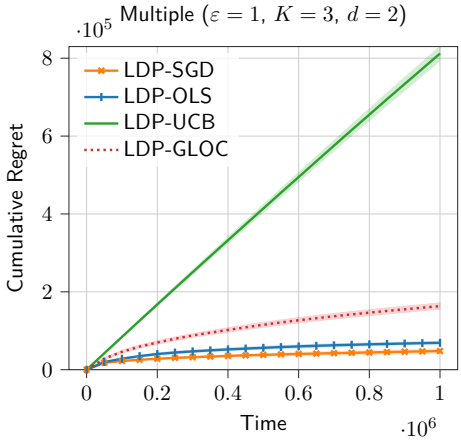
$$Reg(T) \lesssim \left(\frac{K\sigma_{\varepsilon/2,\delta/2}\sqrt{d+\log((TK)/\alpha)}}{\kappa_l p'} + o(1) \right)^{1+\beta} \cdot \begin{cases} \gamma \log T, & \beta = 1, \\ \dfrac{\gamma}{1-\beta} T^{\frac{1-\beta}{2}}, & 0 < \beta < 1. \end{cases}$$

Theorem

Algorithm 2 with SGD mechanism ψ_t^{SGD} and estimator φ_t^{SGD} achieve the following regret with probability at least $1-\alpha$:

$$Reg(T) \lesssim \left(\frac{K r_{\varepsilon,d} L\sqrt{\log((TK\log T)/\alpha)}}{\zeta\kappa_l p'} + o(1) \right)^{1+\beta} \cdot \begin{cases} \gamma L \log T, & \beta = 1, \\ \dfrac{\gamma L}{1-\beta} T^{\frac{1-\beta}{2}}, & 0 < \beta < 1. \end{cases}$$

Experiment – multi-parameter



Thanks!

Reference I

Bastani, H. and M. Bayati (2020). Online decision making with high-dimensional covariates. *Operations Research 68*(1), 276–294.

Chaudhuri, K., C. Monteleoni, and A. D. Sarwate (2011). Differentially private empirical risk minimization. *Journal of Machine Learning Research 12*(3).

Chen, X., D. Simchi-Levi, and Y. Wang (2020). Privacy-preserving dynamic personalized pricing with demand learning. *arXiv*, 1–35.

Ding, B., J. Kulkarni, and S. Yekhanin (2017). Collecting telemetry data privately. *arXiv preprint arXiv:1712.01524*.

Dwork, C. and J. Lei (2009). Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 371–380.

Reference II

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer.

Dwork, C. and A. Roth (2013). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science 9*(3-4), 211–487.

Erlingsson, Ú., V. Pihur, and A. Korolova (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 1054–1067.

Han, Y., Z. Zhou, Z. Zhou, J. Blanchet, P. W. Glynn, and Y. Ye (2020). Sequential batch learning in finite-action linear contextual bandits. *arXiv*.

Kannan, S., J. Morgenstern, A. Roth, B. Waggoner, and Z. S. Wu (2018). A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. *Advances in Neural Information Processing Systems 2018-December*, 2227–2236.

Reference III

Rubinstein, B. I., P. L. Bartlett, L. Huang, and N. Taft (2009). Learning in a large function space: Privacy-preserving mechanisms for svm learning. *arXiv preprint arXiv:0911.5708*.

Shariff, R. and O. Sheffet (2018). Differentially private contextual linear bandits. *Advances in Neural Information Processing Systems 2018-December*, 4296–4306.

Sivakumar, V., S. Wu, and A. Banerjee (2020). Structured linear contextual bandits: A sharp and geometric smoothed analysis. In *International Conference on Machine Learning*, pp. 9026–9035. PMLR.

Smith, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pp. 813–822.

Reference IV

Tang, J., A. Korolova, X. Bai, X. Wang, and X. Wang (2017). Privacy loss in apple's implementation of differential privacy on macos 10.12. *arXiv preprint arXiv:1709.02753*.

Wasserman, L. and S. Zhou (2010). A statistical framework for differential privacy. *Journal of the American Statistical Association 105*(489), 375–389.

Zheng, K., T. Cai, W. Huang, Z. Li, and L. Wang (2020). Locally Differentially Private (Contextual) Bandits Learning. *arXiv* (NeurIPS), 1–20.