A Unified Confidence Sequence for Generalized Linear Models, with Applications to Bandits

Junghyun Lee[1], Se-Young Yun[1], and Kwang-Sung Jun[2]

[1] Kim Jaechul Graduate School of AI, KAIST, [2] Department of Computer Science, University of Arizona

{jh_lee00, yunseyoung}@kaist.ac.kr, kjun@cs.arizona.edu

ICML International Conference On Machine Learning

KAIST AI Kim Jaechul Graduate School

Computer Science

Contributions

- A *unified*, state-of-the-art construction of likelihood ratio-based confidence sequence (CS) for any convex generalized linear models (GLMs), with explicit constants!
- A new CS-based algorithm (`OFUGLB`) that achieves the state-of-the-art regret for self-concordant GLBs.
- Numerical verifications in logistic bandits show the tightness of our new CS and that `OFUGLB` achieves the best numerical regret by a large margin.

Problem Settings

Generalized Linear Models (GLMs)

For a covariate $\boldsymbol{x} \in X$ and an *unknown* parameter $\boldsymbol{\theta}_\star \in \Theta$, the reward r follows the **GLM** if

$$dp(r|\boldsymbol{x}; \boldsymbol{\theta}_\star) = \exp\left(\frac{r\langle \boldsymbol{x}, \boldsymbol{\theta}_\star\rangle - m(\langle \boldsymbol{x}, \boldsymbol{\theta}_\star\rangle)}{g(\tau)} + h(r, \tau)\right) d\nu, \quad (1)$$

where τ is some known scaling (temperature) parameter, and ν is some known base measure (e.g., Lebesgue, counting).

Assumptions:

→ Assumption 1. $X \subseteq \mathcal{B}^d(1)$.

→ Assumption 2. $\boldsymbol{\theta}_\star \in \Theta \subseteq \mathcal{B}^d(S) := \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_2 \leq S\}$ for some known $S > 0$. Also, Θ is nonempty, compact, and convex with intrinsic dimension d.

→ Assumption 3. m is three times differentiable and convex, i.e., m''' exists and $\dot{\mu} := m'' \geq 0$.

Well-known properties:

→ Property 1. $\mathbb{E}[r|\boldsymbol{x}, \boldsymbol{\theta}_\star] = m'(\langle \boldsymbol{x}, \boldsymbol{\theta}_\star\rangle) \triangleq \mu(\langle \boldsymbol{x}, \boldsymbol{\theta}_\star\rangle)$

→ Property 2. $\mathrm{Var}[r|\boldsymbol{x}, \boldsymbol{\theta}_\star] = g(\tau)\dot{\mu}(\langle \boldsymbol{x}, \boldsymbol{\theta}_\star\rangle)$.

μ is the **inverse link (mean) function**.

Question #1

Given a (possibly adaptively-collected) sequential data $\{(x_t, r_t)\}_{t\geq 1}$ sampled from any GLM, output the tightest **confidence sequence (CS)** for $\boldsymbol{\theta}_\star$, i.e., for any $\delta \in (0, 1)$, $\{\mathcal{C}_t(\delta)\}_{t\geq 1}$ such that $\mathbb{P}[\exists t \geq 1 : \boldsymbol{\theta}_\star \notin \mathcal{C}_t(\delta)] \leq \delta$.

Generalized Linear Bandits (GLBs)

First proposed in Filippi et al. [2010] as a nonlinear generalization of linear bandits.

For $t \in [T]$:

1. The learner observes a potentially infinite (contextual) arm-set $\mathcal{X}_t \subset X$
2. The learner chooses $\boldsymbol{x}_t \in \mathcal{X}_t$ according to some policy
3. Receive a reward $r_t|\boldsymbol{x}_t \sim p(\cdot|\boldsymbol{x}_t, \boldsymbol{\theta}_\star)$ (Eqn. (1))

Goal. Minimize:

$$\mathrm{Reg}^B(T) := \sum_{t=1}^{T} \{\mu(\langle \boldsymbol{x}_{t,\star}, \boldsymbol{\theta}_\star\rangle) - \mu(\langle \boldsymbol{x}_t, \boldsymbol{\theta}_\star\rangle)\},$$

where $\boldsymbol{x}_{t,\star} := \arg\max_{\boldsymbol{x}\in\mathcal{X}_t} \mu(\langle \boldsymbol{x}, \boldsymbol{\theta}_\star\rangle)$.

Applications. News recommendations (Bernoulli), social network influence maximization (Poisson), etc [Filippi et al., 2010].

We define the following problem-dependent quantities:

$$\kappa_\star(T) := \left(\frac{1}{T}\sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_{t,\star}^\intercal \boldsymbol{\theta}_\star)\right)^{-1}, \quad \kappa_{\mathcal{X}}(T) := \max_{t\in[T]} \max_{\boldsymbol{x}\in\mathcal{X}_t} \frac{1}{\dot{\mu}(\boldsymbol{x}^\intercal \boldsymbol{\theta}_\star)},$$

$$\text{and} \quad \kappa(T) := \max_{t\in[T]} \max_{\boldsymbol{x}\in\mathcal{X}_t} \max_{\boldsymbol{\theta}\in\Theta} \frac{1}{\dot{\mu}(\boldsymbol{x}^\intercal \boldsymbol{\theta})}.$$

These can scale *exponentially* in S (e.g., Bernoulli)!

$d\sqrt{T/\kappa_\star(T)}$-type regret has been obtained for bounded **GLB**s in a concurrent work of Sawarni et al. [2024], but they make use of explicit warmup and consider limited adaptivity setting.

Question #2

Using our tight CS, how do we obtain tight regret bounds for a wide range of **GLB**s via a *purely optimistic approach*?

A Unified CS for GLMs

We consider log-likelihood-based confidence set "centered" at the *norm-constrained*, batch maximum likelihood estimator (MLE):

$$\mathcal{C}_t(\delta) := \left\{\boldsymbol{\theta} \in \Theta : \mathcal{L}_t(\boldsymbol{\theta}) - \mathcal{L}_t(\widehat{\boldsymbol{\theta}}_t) \leq \beta_t(\delta)^2\right\}, \quad (2)$$

where $\beta_t(\delta)^2$ is the "radius" of the CS that we will define later, and $\mathcal{L}_t(\boldsymbol{\theta})$ is the negative log-likelihood of $\boldsymbol{\theta}$ w.r.t. data collected up to $t-1$, and

$$\mathcal{L}_t(\boldsymbol{\theta}) := \sum_{s=1}^{t-1} \left\{\ell_s(\boldsymbol{\theta}) \triangleq \frac{-r_s\langle \boldsymbol{x}_s, \boldsymbol{\theta}\rangle + m(\langle \boldsymbol{x}_s, \boldsymbol{\theta}\rangle)}{g(\tau)}\right\}, \quad (3)$$

$$\widehat{\boldsymbol{\theta}}_t := \arg\min_{\boldsymbol{\theta}\in\Theta} \mathcal{L}_t(\boldsymbol{\theta}). \quad (4)$$

Theorem 3.1. Let $L_t := \max_{\boldsymbol{\theta}\in\Theta} \|\nabla\mathcal{L}_t(\boldsymbol{\theta})\|_2$ be the Lipschitz constant of $\mathcal{L}_t(\cdot)$ that may depend on $\{(\boldsymbol{x}_s, r_s)\}_{s=1}^{t-1}$. Then, we have $\mathbb{P}[\exists t \geq 1 : \boldsymbol{\theta}_\star \notin \mathcal{C}_t(\delta)] \leq \delta$, where

$$\mathcal{C}_t(\delta) = \left\{\boldsymbol{\theta} \in \Theta : \mathcal{L}_t(\boldsymbol{\theta}) - \mathcal{L}_t(\widehat{\boldsymbol{\theta}}_t) \leq \beta_t(\delta)^2\right\}, \quad (5)$$

where $\beta_t(\delta)^2 \leq \log\frac{1}{\delta} + d\log\left(e \vee \frac{2eSL_t}{d}\right)$.

For Bernoulli, our radius of $\mathcal{O}_\delta(d\log(St/d))$ this is a strict improvement over prior $\mathcal{O}_\delta\left(d\log(St/d) + S\right)$ of Lee et al. [2024].

→ Remark. *This resolves an open problem posited by Lee et al. [2024] on* poly(S)*-free CS for Bernoulli.*

We consider the following additional assumption on the GLM:

→ Assumption 4. (self-concordance) For some $R_s \in (0, \infty)$, $|\ddot{\mu}(\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle)| \leq R_s\dot{\mu}(\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle)$ for all $\boldsymbol{x} \in X, \boldsymbol{\theta} \in \Theta$.

For this class of GLMs, we have a slightly relaxed *ellipsoidal* CS:

Theorem 3.2. With the same notations as Theorem 3.1, we have $\mathbb{P}[\exists t \geq: \boldsymbol{\theta}_\star \notin \mathcal{E}_t(\delta)] \leq \delta$, where

$$\mathcal{E}_t(\delta) := \left\{\boldsymbol{\theta} \in \Theta : \left\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\right\|^2_{\nabla^2\mathcal{L}_t(\widehat{\boldsymbol{\theta}}_t) + \frac{1+SR_s}{2S^2}\boldsymbol{I}_d} \leq \gamma_t(\delta)^2\right\}, \quad (6)$$

where $\gamma_t(\delta)^2 := 2(1 + SR_s)(1 + \beta_t(\delta)^2)$.

→ Remark. *This is easier to implement in practice, and for bandits, this amounts to a closed-form bonus in UCB.*

Proof via PAC-Bayes with Uniform Prior/Posterior

1. PAC-Bayesian Time-Uniform Bound.

Lemma 3.3. For any data-independent prior $\mathbb{Q}$ and any sequence of adapted posterior distributions $\{\mathbb{P}_t\}$, the following holds: for any $\delta \in (0, 1)$,

$$\mathbb{P}\left(\exists t \geq 1 : \mathcal{L}_t(\boldsymbol{\theta}_\star) - \mathbb{E}_{\boldsymbol{\theta}\sim\mathbb{P}_t}[\mathcal{L}_t(\boldsymbol{\theta})] \geq \log\frac{1}{\delta} + \mathrm{KL}(\mathbb{P}_t||\mathbb{Q})\right) \leq \delta. \quad (7)$$

Proof sketch. This is a standard recipe using Ville's inequality and Donsker-Varadhan variational representation of KL; see Chugg et al. [2023] for relevant references.

2. Novel choice of $\mathbb{Q}$ and $\mathbb{P}_t$.

For $c \in (0, 1]$ to be determined later, we set

$$\mathbb{Q} = \mathrm{Unif}(\Theta), \quad \mathbb{P}_t = \mathrm{Unif}(\widetilde{\Theta}_t \triangleq (1-c)\widehat{\boldsymbol{\theta}}_t + c\Theta), \quad (8)$$

where $\boldsymbol{a} + \Theta = \{\boldsymbol{a} + \boldsymbol{\theta} : \boldsymbol{\theta} \in \Theta\}$ for a vector $\boldsymbol{a} \in \mathbb{R}^d$. Then, we have

$$\mathrm{KL}(\mathbb{P}_t||\mathbb{Q}) = \log\frac{\mathrm{vol}(\Theta)}{\mathrm{vol}(\widetilde{\Theta})} = d\log\frac{1}{c}.$$

3. Lipschitzness of $\mathcal{L}_t(\cdot)$.

We also have that

$$\mathbb{E}_{\boldsymbol{\theta}\sim\mathbb{P}_t}[\mathcal{L}_t(\boldsymbol{\theta})] = \mathcal{L}_t(\widehat{\boldsymbol{\theta}}_t) + \mathbb{E}_{\boldsymbol{\theta}\sim\mathbb{P}_t}[\mathcal{L}_t(\boldsymbol{\theta}) - \mathcal{L}_t(\widehat{\boldsymbol{\theta}}_t)] \leq \mathcal{L}_t(\widehat{\boldsymbol{\theta}}_t) + 2SL_tc,$$

where the last inequality follows from the Lipschitzness of $\mathcal{L}_t(\cdot)$ and the observation that for $\boldsymbol{\theta} = (1-c)\widehat{\boldsymbol{\theta}}_t + c\widetilde{\boldsymbol{\theta}} \in \widetilde{\Theta}_t$ and $\left\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\right\|_2 = c\left\|\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_t\right\|_2 \leq 2Sc$. We conclude by choosing minimizing over $c \in (0, 1]$. The expression in Theorem 3.1 follows from $c = 1 \wedge \frac{d}{2SL_t}$. □

→ Remark. *Such choices of $\mathbb{Q}$ and $\mathbb{P}_t$ have been considered previously in universal portfolios [Blum and Kalai, 1999] and fast rates in online learning [Foster et al., 2018]. This is the first time such a translated/shrunken posterior has been used in the PAC-Bayes context.*

`OFUGLB`

`OFUGLB` is of the following form:

1. Obtain $\widehat{\boldsymbol{\theta}}_t$ (Eqn. (4)) and $\mathcal{C}_t(\delta)$ (Theorem 3.1)
2. Solve $(\boldsymbol{x}_t, \boldsymbol{\theta}_t) = \arg\max_{\boldsymbol{x}\in\mathcal{X}_t, \boldsymbol{\theta}\in\mathcal{C}_t(\delta)} \mu(\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle)$
3. Play $\boldsymbol{x}_t$, then observe/receive a reward $r_t \in \{0, 1\}$.

We then have the following *state-of-the-art* regret bound:

Theorem 4.1. `OFUGLB` attains the following regret bound with probability at least $1 - \delta$:

$$\mathrm{Reg}^B(T) \lesssim_\delta d\sqrt{\frac{g(\tau)T}{\kappa_\star(T)}} + d^2 R_S R_{\dot{\mu}}\sqrt{g(\tau)}\kappa(T),$$

where $R_{\dot{\mu}} := \max_{\boldsymbol{x}\in\mathcal{X}_{[T]}, \boldsymbol{\theta}\in\Theta} \dot{\mu}(\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle)$.

→ Remark. Nontrivial technical contributions, including a new optimistic upper bound of regret, self-concordant control, etc.

Linear bandits. $\widetilde{\mathcal{O}}(\sigma d\sqrt{T})$

→ matches prior state-of-the-art [Flynn et al., 2023]

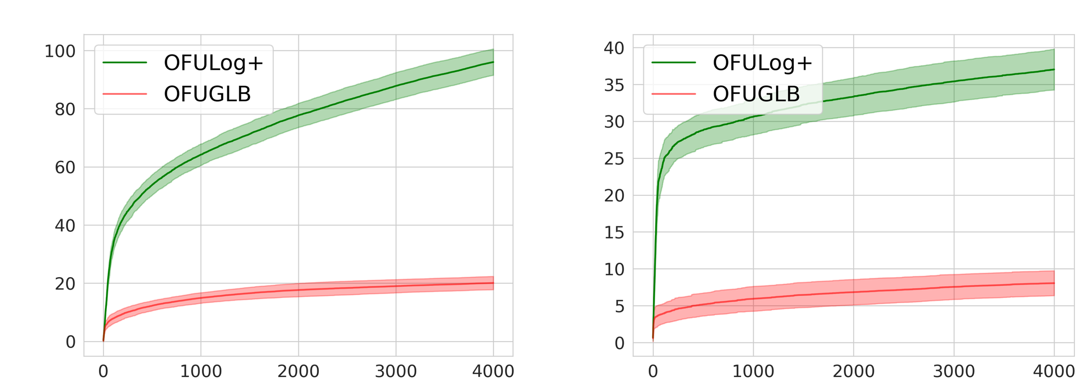
Logistic bandits. $\widetilde{\mathcal{O}}(d\sqrt{T/\kappa_\star(T)} + d^2\kappa(T))$

→ first poly(S)-free regret with *purely optimistic approach*, improves upon `OFULog+` of Lee et al. [2024]!

Poisson bandits. $\widetilde{\mathcal{O}}(dS\sqrt{T/\kappa_\star(T)} + d^2e^{2S}\kappa(T))$

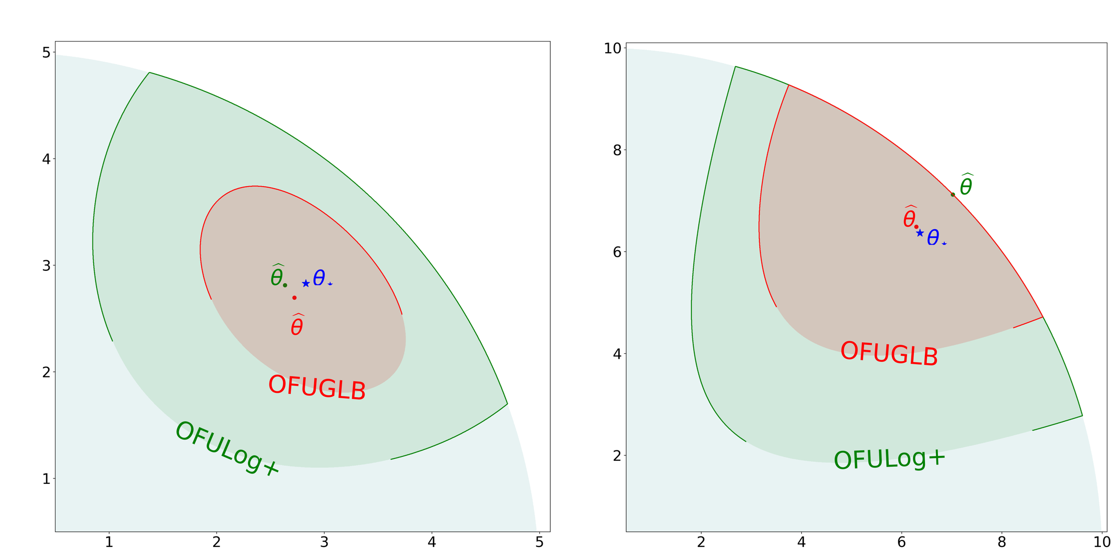
→ first regret guarantee!

Experiments for Logistic Bandits



(a) $S = 5$ (b) $S = 10$

Figure 1: Numerical regrets.



(a) $S = 5$ (b) $S = 10$

Figure 2: Confidence sets at $t = 4000$ from a single run.

Future Directions

- Extension to kernelized/functional GLMs?
- Implications to RLHF; see e.g., Das et al. [2024].
- Arm-set geometry-dependent transient term for **GLB**s
- Regret lower bound of general **GLB**s

References

A. Blum and A. Kalai. Universal Portfolios With and Without Transaction Costs. *Machine Learning*, 35(3):193–205, 1999.

B. Chugg et al. A Unified Recipe for Deriving (Time-Uniform) PAC-Bayes Bounds. *Journal of Machine Learning Research*, 24(372):1–61, 2023.

N. Das et al. Provably Sample Efficient RLHF via Active Preference Optimization. *arXiv preprint arXiv:2402.10500*, 2024.

S. Filippi et al. Parametric Bandits: The Generalized Linear Case. In *NIPS*, 2010.

H. Flynn et al. Improved Algorithms for Stochastic Linear Bandits Using Tail Bounds for Martingale Mixtures. In *NeurIPS*, 2023.

D. J. Foster et al. Logistic Regression: The Importance of Being Improper. In *COLT*, 2018.

J. Lee et al. Improved Regret Bounds of (Multinomial) Logistic Bandits via Regret-to-Confidence-Set Conversion. In *AISTATS*, 2024.

A. Sawarni et al. Optimal Regret with Limited Adaptivity for Generalized Linear Contextual Bandits. *arXiv preprint arXiv:2404.06831*, 2024.