Improved Confidence Bounds for the Linear Logistic Model and Applications to Bandits




Kwang-Sung Jun
UArizona
(presenter)

Lalit Jain
UW

Blake Mason
UW-Madison

Houssam Nassif
Amazon

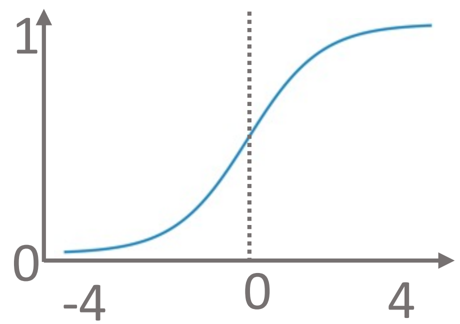
Logistic linear bandits

- Given an arm set $\mathcal{X} \subset \mathbb{R}^d$
- For $t = 1 \dots T$
  - A random user u arrives.
  - The system chooses an arm $x_t \in \mathcal{X}$ and shows it.
  - The user u provides a reward $y_t \in \{0,1\}$ on the item x_t .

[Assumption] $y_t \sim \text{Bernoulli}(\, \mu(x_t^\top \theta^*)\,)$

unknown parameter

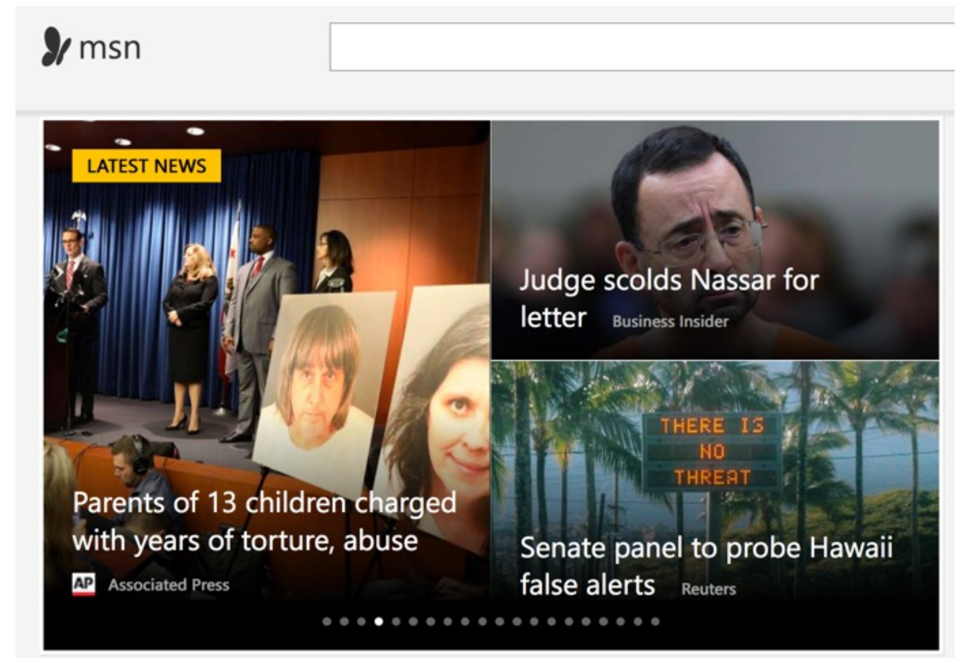
$\mu(z) = 1/(1 + \exp(-z))$



Two different settings

- (Reward maximization) maximize cumulative reward: $\sum_{t=1}^{T} y_t$
- (Pure exploration) identify the best arm $x^* = \arg\max_{x \in \mathcal{X}} \mu(x^\top \theta^*)$

Example: online news recommendation



Contribution 1: Prediction error bound

Fix $x \in \mathbb{R}^d$. With probability at least $1-\delta$,

$$H_t(\theta^*) := \sum_{s=1}^{t} \dot{\mu}(x_s^\top \theta^*) x_s x_s^\top$$

$$\underbrace{\max_{s=1..t} \|x_s\|^2_{H_t(\theta^*)^{-1}} \lesssim \frac{1}{d}}_{\text{warmup condition}} \implies \Big|\underbrace{x^\top \hat{\theta}_t}_{\text{predicted}} - \underbrace{x^\top \theta^*}_{\text{true}}\Big| \le 2.4\sqrt{\underbrace{\boxed{x^\top H_t(\theta^*)^{-1} x}}_{\text{variance}} \cdot \log(1/\delta)}$$

MLE

A **finite-time bound** that is **asymptotically optimal** for the first time (order-wise).

- Improved upon the prior art Li et al. (2017)
- Valid once we have d^2 samples (prior art: d^3)
- No explicit dependence on the infamous $\kappa^{-1} \approx \exp(\|\theta^*\|_2)$

Contribution 2: Pure exploration

- Our sample complexity bound

$\Delta_{\mathcal{X}}$:= a probability distribution over $\mathcal{X}$

$$\Bigg(\underbrace{d^2\kappa^{-1}}_{\text{warmup}} + \underbrace{\min_{\lambda\in\Delta_{\mathcal{X}}} \max_{x\in\mathcal{X}\setminus\{x^*\}} \frac{\|x^*-x\|^2_{H(\lambda,\theta^*)^{-1}}}{\left((x^*-x)^\top\theta^*\right)^2}}_{\text{the main term (instance-dependent)}}\Bigg)\log\left(\frac{|\mathcal{X}|}{\delta}\right) \lesssim \left(d^2\kappa^{-1} + \frac{d\kappa^{-1}}{\Delta_{\min}^2}\right)\log\left(\frac{|\mathcal{X}|}{\delta}\right)$$

$\Delta_{\min} := \min_{x\in\mathcal{X}\setminus\{x^*\}} (x^*-x)^\top\theta^*$

- Kazerouni et al. (2019): $\frac{d\kappa^{-2}|\mathcal{X}|}{\Delta_{\min}^2}\log\left(\frac{1}{\delta}\right)$

- In fact, our bound works for a more general setting called **transductive** pure exploration.
- Optimality
  - [Warmup] (thm) It is impossible to avoid the dependence on κ^{-1} in the worst case.
  - [Main term] Not exactly tight, but quite close (Taylor approximation)

Contribution 3: K-armed contextual bandits

- Arm set $\mathcal{X}_t$ is changing with t; $K \coloneqq \max_{t=1..T} |\mathcal{X}_t|$
- Stochastic context assumption (following Li et al. (2017))

Expected regret bound

- **Ours**: $O(\sqrt{dT \log K} + d^3\kappa^{-1} + d^5)$
- **Li et al. (2017)**: $O(\boxed{\kappa^{-1}}\sqrt{dT \log K} + d^3\boxed{\kappa^{-4}} + d^5)$
  - Strictly worse than ours.
- **Faury et al. (2020)**: $O\left(\boxed{d}\sqrt{T} + d^2\kappa^{-1}\right)$
  - Worse than ours when both d and T are large
  - Better than ours when $K = \Omega(e^d)$

: key difference

* the first two bounds are based on the best-case context distribution