Choosing Answers in ε-Best-Answer Identification for Linear Bandits

Marc Jourdan and Rémy Degenne

June 17, 2022

Motivation

Initial goal: Identify the item having the highest averaged return.

Problem: When the two best items have highly similar averaged return, the number of samples required to differentiate them is large.

Corrected goal: Identify one item which is ε-close to the best one (ε-BAI).

Challenge: Multiple correct answers.

? How to choose among the set of ε-optimal answers ?

☞ Focus on the ε-optimal answer which is the easiest to verify.

ε-BAI for Transductive linear Gaussian bandits

Transductive linear Gaussian bandits:

- arm $a \in \mathcal{K}$, finite subset of $\mathbb{R}^d$,
- answer $z \in \mathcal{Z}$, finite subset of $\mathbb{R}^d$,
- unknown mean parameter, $\mu \in \mathbb{R}^d$.

At time t, pull $a_t \in \mathcal{K}$ and observe $X_t^{a_t} \sim \mathcal{N}\left(\langle \mu, a_t \rangle, 1\right)$.

Goal: Identify one ε-optimal answer with confidence δ, $z \in \mathcal{Z}_\varepsilon(\mu)$.

Objective: Minimize $\mathbb{E}_\mu[\tau_\delta]$ for (ε, δ)-PAC algorithms

$$\mathbb{P}_\mu\left[\tau_\delta < +\infty,\ z_{\tau_\delta} \notin \mathcal{Z}_\varepsilon(\mu)\right] \leq \delta \,.$$

? What is the best one could achieve ? Degenne and Koolen (2019)

☞ For all (ε, δ)-PAC strategy, for all μ, $\liminf_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\ln(1/\delta)} \geq T_\varepsilon(\mu)$.

Furthest answer

? How to choose among the set of ε-optimal answers ?

☞ **Furthest answer**: ε-optimal answer for which its alternative is the easiest to differentiate from thanks to an optimal allocation over arms.

$$(z_F(\mu), w_F(\mu)) \stackrel{\text{def}}{=} \arg\max_{(z,w)\in\mathcal{Z}_\varepsilon(\mu)\times\triangle_K} \inf_{\lambda\in\neg_\varepsilon z} \frac{1}{2}\|\mu-\lambda\|^2_{V_w} \, ,$$

are the maximizers realizing $T_\varepsilon(\mu)$. $\neg_\varepsilon z$ alternative to z, $\triangle_K$ simplex, $V_w = \sum_{a\in\mathcal{K}} w^a a a^\intercal$ design matrix with norm $\|\cdot\|_{V_w}$.

Greedy answer: $z^\star(\mu) = \arg\max_{z\in\mathcal{Z}}\langle\mu, z\rangle$, unique correct answer in BAI.

☞ sample inefficient, 10% higher empirical stopping time for $\delta = 1\%$.

Adapting any BAI algorithm for ε-BAI

? How to stop to obtain an (ε, δ)-PAC strategy ?

☞ calibrated **GLR stopping rule** for $z_t \in \mathcal{Z}_\varepsilon(\mu_{t-1})$

? Which z_t should we **recommend** to stop as early as possible ?

☞ Instantaneous furthest answer: ε-optimal answer with highest GLR

$$z_F(\mu_{t-1}, N_{t-1}) = \arg\max_{z \in \mathcal{Z}_\varepsilon(\mu_{t-1})} \inf_{\lambda \in \neg_\varepsilon z_t} \|\mu_{t-1} - \lambda\|^2_{V_{N_{t-1}}} ,$$

where $N^a_{t-1} = \sum_{s=1}^{t-1} \mathbf{1}_{\{a_s = a\}}$ and $\mu_{t-1} = V^{-1}_{N_{t-1}} \sum_{s=1}^{t-1} X^{a_s}_s a_s$.

? How to **modify any BAI algorithms to be** (ε, δ)**-PAC** ?

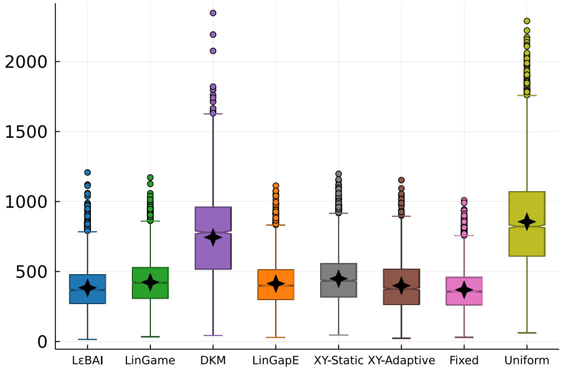
☞ use GLR stopping with $z_t \in z_F(\mu_{t-1}, N_{t-1})$,

☞ keep the sampling rule unchanged.

LεBAI (Linear ε-BAI)

Can we achieve asymptotic optimality and be empirically competitive ?

☞ LεBAI, by using the concept of furthest answer in the sampling rule.



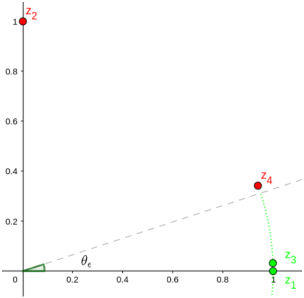


Figure: Empirical stopping time of LεBAI compared to modified BAI algorithms.

Conclusion

Paper & Poster

1. Don't choose greedily: aim at identifying the *furthest* answer !
2. Simple procedure to adapt your favorite BAI algorithm to ε-BAI.
3. LεBAI, asymptotically optimal and empirically competitive.