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

Marc Jourdan, Rémy Degenne
Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9198-CRIStAL, F-59000 Lille, France

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.

Problem Statement

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 bounded mean parameter, $\mu \in \mathcal{M} \subseteq \mathbb{R}^d$.

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

Goal: Identify one ε-optimal answer, $z \in \mathcal{Z}_\varepsilon(\mu)$ with $\varepsilon \geq 0$.

Two notions of ε-optimality:
- additive, $\mathcal{Z}_\varepsilon^{\text{add}}(\mu) = \{z \in \mathcal{Z} : \langle \mu, z \rangle \geq \max_{z \in \mathcal{Z}} \langle \mu, z \rangle - \varepsilon\}$,
- multiplicative, $\mathcal{Z}_\varepsilon^{\text{mul}}(\mu) = \{z \in \mathcal{Z} : \langle \mu, z \rangle \geq (1-\varepsilon) \max_{z \in \mathcal{Z}} \langle \mu, z \rangle\}$.

Greedy answer, $z^\star(\mu) = \arg\max_{z \in \mathcal{Z}} \langle \mu, z \rangle$, unique correct answer in BAI ($\varepsilon = 0$).

(ε, δ)-PAC identification strategy

Fixed-confidence setting, $\delta \in (0,1)$. Three rules:
- *sampling* rule, $a_t \in \mathcal{K}$,
- *recommendation* rule, $z_t \in \mathcal{Z}$,
- *stopping* rule, τ_δ.

Requirement: (ε, δ)-PAC, $\mathbb{P}_\mu [\tau_\delta < +\infty, z_{\tau_\delta} \notin \mathcal{Z}_\varepsilon(\mu)] \leq \delta$.

Objective: Minimize $\mathbb{E}_\mu[\tau_\delta]$.

? What is the best one could achieve ?

☞ Degenne and Koolen (2019): For all (ε, δ)-PAC strategy, for all $\mu \in \mathcal{M}$,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\ln(1/\delta)} \geq T_\varepsilon(\mu) \, ,$$

where the inverse of the characteristic time is

$$T_\varepsilon(\mu)^{-1} = \max_{z \in \mathcal{Z}_\varepsilon(\mu)} \max_{w \in \triangle_K} \inf_{\lambda \in \neg_\varepsilon z} \frac{1}{2} \|\mu - \lambda\|_{V_w}^2 \, .$$

Alternative to $z \in \mathcal{Z}$: $\neg_\varepsilon z = \overline{\{\lambda \in \mathcal{M} : z \notin \mathcal{Z}_\varepsilon(\lambda)\}}$.

$\triangle_K$ simplex, $V_w = \sum_{a \in \mathcal{K}} w^a a a^\intercal$ design matrix with norm $\| \cdot \|_{V_w}$.

Furthest answer

Identifying z as an ε-optimal answer is equivalent to rejecting its alternative.

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

☞ **Furthest answer:** $z_F(\mu)$ is the ε-optimal answer for which its alternative is the easiest to reject by using an optimal allocation over arms $w_F(\mu)$.

$$(z_F(\mu), w_F(\mu)) = \operatorname*{arg\,max}_{(z,w) \in \mathcal{Z}_\varepsilon(\mu) \times \triangle_K} \inf_{\lambda \in \neg_\varepsilon z} \frac{1}{2} \|\mu - \lambda\|_{V_w}^2 \, .$$

Assumption: unique furthest answer, i.e. $|z_F(\mu)| = 1$.

Numerical simulations: $z_1 = \mu = (1,0)$, $z_2 \in \mathcal{Z}_\varepsilon^{\text{mul}}(\mu)$ and $z_3, z_4 \in \mathcal{Z} \setminus \mathcal{Z}_\varepsilon^{\text{mul}}(\mu)$.

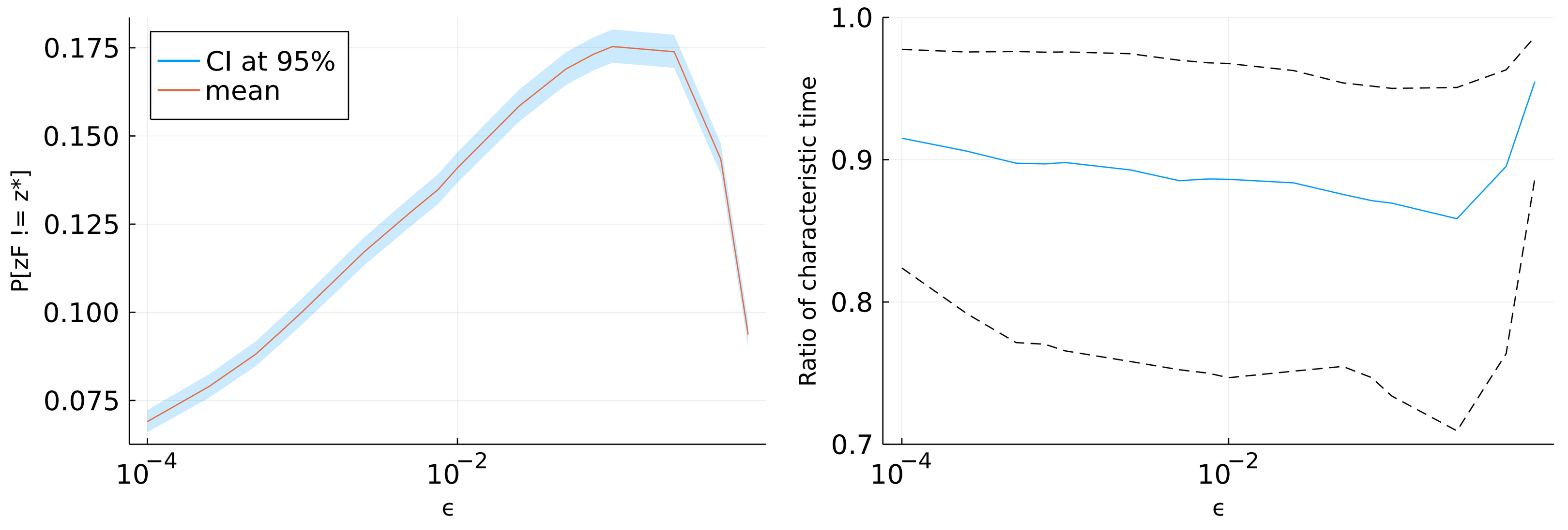

Figure 1: (a) Proportion of $z_F(\mu) \notin z^\star(\mu)$. (b) Ratio between $T_\varepsilon^{\text{mul}}(\mu)$ and the value at $z^\star(\mu)$.

Adapting any BAI algorithm for ε-BAI

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

☞ **GLR stopping rule:** Given $z_t \in \mathcal{Z}_\varepsilon(\mu_{t-1})$, stop when

$$\inf_{\lambda \in \neg_\varepsilon z_t} \|\mu_{t-1} - \lambda\|_{V_{N_{t-1}}}^2 > 2\beta(t-1, \delta) \, , \tag{1}$$

where $N_{t-1}^a = \sum_{s=1}^{t-1} \mathbf{1}_{\{a_s = a\}}$, $\mu_{t-1} = V_{N_{t-1}}^{-1} \sum_{s=1}^{t-1} X_s^{a_s} a_s$ and

$$\beta(t, \delta) = 2K \ln(4 + \ln(t/K)) + K \mathcal{C}^{gG}(\ln(1/\delta)/K) \, , \tag{2}$$

with $\mathcal{C}^{gG}(x) \approx x + \ln(x)$, see Kaufmann and Koolen (2018).

? Which $z_t \in \mathcal{Z}_\varepsilon(\mu_{t-1})$ should we **recommend** to stop as early as possible ?

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

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

Other choices are inefficient: greedy (samples) or furthest (computation) answers.

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

☞ use GLR stopping rule with $z_t \in z_F(\mu_{t-1}, N_{t-1})$,
☞ keep the sampling rule unchanged.

10% lower empirical stopping time when using $z_F(\mu_{t-1}, N_{t-1})$ instead of $z^\star(\mu_{t-1})$.

LεBAI

Input: $\mathcal{Z}$-oracle $\mathcal{L}^{\mathcal{Z}}$ and learner $\mathcal{L}^{\mathcal{K}}$ on $\triangle_K$.
Pull once each arm $a \in \mathcal{K}$, set $n_0 = K$ and $W_{n_0} = 1_K$;

For $t \geq n_0 + 1$

- Get $z_t \in z_F(\mu_{t-1}, N_{t-1})$;
- If (1) holds for z_t then return z_t;
- Get $\left(\tilde{z}_t, w_t^{\mathcal{L}^{\mathcal{K}}}\right)$ from $\mathcal{L}^{\mathcal{Z}} \times \mathcal{L}^{\mathcal{K}}$;
- Let $w_t = \frac{1_K}{tK} + \left(1 - \frac{1}{t}\right) w_t^{\mathcal{L}^{\mathcal{K}}}$ and $W_t = W_{t-1} + w_t$;
- Closest alternative: $\lambda_t \in \arg\min_{\lambda \in \neg_\varepsilon \tilde{z}_t} \|\mu_{t-1} - \lambda\|_{V_{w_t}}^2$;
- Optimistic gains: $\forall a \in \mathcal{K}$, $U_t^a = \left(\|\mu_{t-1} - \lambda_t\|_{aa^\intercal} + \sqrt{c_{t-1}^a}\right)^2$;
- Feed $\mathcal{L}^{\mathcal{K}}$ with gain $g_t(w) = (1 - \frac{1}{t})\langle w, U_t \rangle$;
- Pull $a_t \in \arg\min_{a \in \mathcal{K}} N_{t-1}^a - W_t^a$, observe $X_t^{a_t}$;

Theorem 1. *Let $\mathcal{L}^{\mathcal{K}}$ with sub-linear regret (e.g. AdaHedge) and $\mathcal{L}^{\mathcal{Z}}$ returning $\tilde{z}_t \in z_F(\mu_{t-1})$. Using (2) as stopping threshold $\beta(t, \delta)$, LεBAI yields an (ε, δ)-PAC algorithm and, for all $\mu \in \mathcal{M}$ such that $|z_F(\mu)| = 1$,*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\ln(1/\delta)} \leq T_\varepsilon(\mu) \, .$$

Efficient heuristic: $\mathcal{L}^{\mathcal{Z}}$ uses $\tilde{z}_t = z_t$.

Experiments

Hard instance with $\mathcal{K} = \mathcal{Z}$: $z_1 = \mu = (1,0)$, $z_2 = (0,1)$, $z_3 = (\cos(\phi_1), \sin(\phi_1))$, $z_4 = (\cos(\phi_2), \sin(\phi_2))$ where $(\phi_1, \phi_2) = \left(\frac{1}{10}\theta_\varepsilon, \frac{11}{10}\theta_\varepsilon\right)$, $\theta_\varepsilon = \arccos(1 - \varepsilon)$ and $\varepsilon = 5\%$.

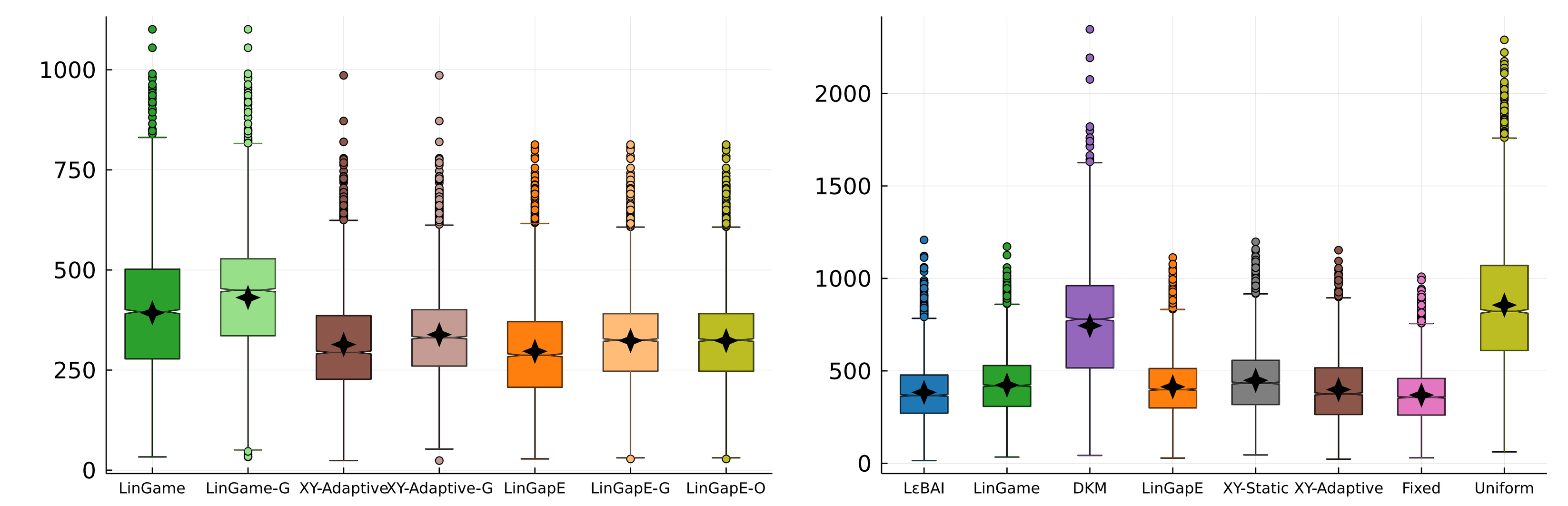

Figure 2: Empirical stopping time at $\delta = 1\%$ (star equals mean) for **(a)** modified BAI algorithms (add) and **(b)** heuristic LεBAI (mul). "-G" is $z_t \in z^\star(\mu_{t-1})$. "-O" is the ε-gap stopping rule with $z_t \in z^\star(\mu_{t-1})$.

Conclusion

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.