An ε-Best-Arm Identification Algorithm for Fixed-Confidence and Beyond

Marc Jourdan, Rémy Degenne and Émilie Kaufmann

November 3, 2023

Centre de Recherche en Informatique, Signal et Automatique de Lille

Motivation

Goal: Identify one item that has a good enough average return.

Two main approaches:

- **fixed-confidence**, control the error and minimize the sampling budget or
- **fixed-budget**, control the sampling budget and minimize the error.

⚠ Too restrictive for many applications !

☞ This paper: **guarantees at any time** on the candidate answer !

ε-Best-arm identification (ε-BAI)

K arms: arm $i \in [K]$ with ν_i is a 1-sub-Gaussian with mean μ_i.

Goal: identify one ε-good arm $\mathcal{I}_\varepsilon(\mu) = \{i \mid \mu_i \geq \max_j \mu_j - \varepsilon\}$.

Algorithm: at time n,

- *Recommendation rule:* recommend the candidate answer $\hat{\imath}_n$
- **Sampling rule**: pull arm I_n and observe $X_n \sim \nu_{I_n}$.

Fixed-confidence: given an error/confidence pair (ε, δ), define an (ε, δ)-PAC stopping time $\tau_{\varepsilon,\delta}$, i.e. $\mathbb{P}_\nu(\tau_{\varepsilon,\delta} < +\infty, \hat{\imath}_{\tau_{\varepsilon,\delta}} \notin \mathcal{I}_\varepsilon(\mu)) \leq \delta$,
☞ Minimize the **expected sample complexity** $\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]$.

Fixed-budget: given an error/budget pair (ε, T),
☞ Minimize the **probablity of ε-error** $\mathbb{P}_\nu(\hat{\imath}_T \notin \mathcal{I}_\varepsilon(\mu))$ at time T.

Anytime: Control the **simple regret** $\mathbb{E}_\nu[\max_j \mu_j - \mu_{\hat{\imath}_n}]$ at any time n.

Top Two sampling rule: EB-TC$_{\varepsilon_0}$, **fixed** β or **IDS**

Input: **slack** $\varepsilon_0 > 0$, proportion $\beta \in (0,1)$ (only for fixed).

Set $\hat{\imath}_n \in \arg\max_{i \in [K]} \mu_{n,i}$;

Set $B_n = \hat{\imath}_n$ and $C_n \in \arg\min_{i \neq B_n} \frac{\mu_{n,B_n} - \mu_{n,i} + \varepsilon_0}{\sqrt{1/N_{n,B_n} + 1/N_{n,i}}}$;

Update $\bar{\beta}_{n+1}(B_n, C_n)$ where $\beta_n(i,j) = \begin{cases} \beta & \textbf{[fixed]} \\ \frac{N_{n,j}}{N_{n,i}+N_{n,j}} & \textbf{[IDS]} \end{cases}$;

Tracking: $I_n = \begin{cases} C_n & \text{if } N^{B_n}_{n,C_n} \le (1 - \bar{\beta}_{n+1}(B_n, C_n)) T_{n+1}(B_n, C_n) \\ B_n & \text{otherwise} \end{cases}$;

Output: next arm to sample I_n and next recommendation $\hat{\imath}_n$.

$(N_{n,i}, \mu_{n,i})$: pulling count and empirical mean of arm i before time n.
$T_n(i,j)$: selection count of the leader/challenger pair (i,j) before time n.
$\bar{\beta}_n(i,j)$: average proportion when selecting (i,j) before time n.
$N^i_{n,j}$: pulling count of arm j when selecting pair (i,j) before time n.

Fixed-confidence guarantees

(Degenne and Koolen, 2019) For all (ε, δ)-PAC algorithms,

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

GLR$_\varepsilon$ stopping rule: recommend $\hat{\imath}_n \in \arg\max_{i \in [K]} \mu_{n,i}$ and

$$\tau_{\varepsilon,\delta} = \inf\{n > K \mid \min_{i \neq \hat{\imath}_n} \frac{\mu_{n,\hat{\imath}_n} - \mu_{n,i} + \varepsilon}{\sqrt{1/N_{n,\hat{\imath}_n} + 1/N_{n,i}}} \geq \sqrt{2c(n-1,\delta)}\} \, . \qquad (1)$$

Theorem

Let $\varepsilon > 0$. Combined with GLR$_\varepsilon$ stopping (1), EB-TC$_\varepsilon$ with IDS (resp. fixed β) proportions is **asymptotically** *(resp. β-)***optimal** *in fixed-confidence ε-BAI for Gaussian distributions.*

☞ EB-TC$_\varepsilon$ has also guarantees for **any confidence level**.

Anytime guarantees

Probability of ε-error and **expected simple regret**.

Theorem

Let $\varepsilon_0 > 0$. EB-TC$_{\varepsilon_0}$ with fixed $\beta = 1/2$ satisfies that, for all $n > 5K^2/2$,

$$\forall \varepsilon \geq 0, \quad \mathbb{P}_\nu \left(\hat{\imath}_n \notin \mathcal{I}_\varepsilon(\mu) \right) \leq \exp\left(-\Theta\left(\frac{n}{H_{i_\mu(\varepsilon)}(\mu, \varepsilon_0)} \right) \right) ,$$

$$\mathbb{E}_\nu[\mu_\star - \mu_{\hat{\imath}_n}] \leq \sum_{i \in [C_\mu - 1]} (\Delta_{i+1} - \Delta_i) \exp\left(-\Theta\left(\frac{n}{H_i(\mu, \varepsilon_0)} \right) \right) ,$$

where $H_1(\mu, \varepsilon_0) = K(2\Delta_{\min}^{-1} + 3\varepsilon_0^{-1})^2$ and $H_i(\mu, \varepsilon_0) = \Theta(K/\Delta_{i+1}^{-2})$.

Ordered distinct gaps $(\Delta_i)_{i \in [C_\mu]}$ and $i_\mu(\varepsilon) = i$ if $\varepsilon \in [\Delta_i, \Delta_{i+1})$.

Empirical results

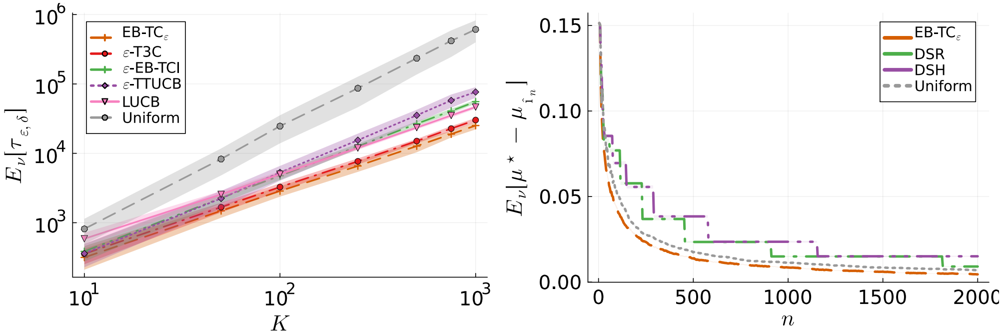

Figure: (a) Stopping time on instances $\mu_i = 1 - ((i-1)/(K-1))^{0.3}$ for varying K. (b) Simple regret on instance $\mu = (0.6, 0.6, 0.55, 0.45, 0.3, 0.2)$ fors EB-TC$_{\varepsilon_0}$ with slack $\varepsilon_0 = 0.1$ and fixed $\beta = 1/2$.

GLR$_\varepsilon$ stopping (1) with $(\varepsilon, \delta) = (10^{-1}, 10^{-2})$. T3C, EB-TCI, TTUCB, TaS, FWS, DKM are modified for ε-BAI.

Conclusion

Benefits of EB-TC$_\varepsilon$:

1. Easy to implement, computationally inexpensive and versatile algorithm.
2. Good empirical performance for the sample complexity and simple regret.
3. Asymptotic and finite confidence upper bound on the expected sample complexity.
4. Anytime upper bounds on the uniform ε-error and the expected simple regret.

Paper & Poster