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

Marc Jourdan, Rémy Degenne and Emilie Kaufmann
Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189-CRIStAL, F-59000 Lille, France

Motivation

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

Two main approaches:

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

⚠ Too restrictive for many applications !

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

ε-Best-arm identification (ε-BAI)

K arms: $\nu_i \in \mathcal{D}$ is the 1-sub-Gaussian distribution of arm $i \in [K]$ with mean μ_i.

Goal: identify one of the ε-good arms $\mathcal{I}_\varepsilon(\mu) = \{i \mid \mu_i \geq \mu_\star - \varepsilon\}$ with $\mu^\star = \max_i \mu_i$.

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 $(\varepsilon, \delta) \in \mathbb{R}_+ \times (0,1)$, define a stopping time $\tau_{\varepsilon,\delta}$ which is (ε,δ)-PAC, i.e. $\mathbb{P}_\nu(\tau_{\varepsilon,\delta} < +\infty, \hat{\imath}_{\tau_{\varepsilon,\delta}} \notin \mathcal{I}_\varepsilon(\mu)) \leq \delta$, and
☞ Minimize the **expected sample complexity** $\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]$.

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

Anytime: Minimize the **expected simple regret** $\mathbb{E}_\nu[\mu^\star - \mu_{\hat{\imath}_n}]$ at any time n.

Lower bound on the expected sample complexity

? What is the best one could achieve ?

☞ Degenne and Koolen (2019): For all (ε,δ)-PAC algorithms and all Gaussian instances with $\mu \in \mathbb{R}^K$, $\liminf_{\delta\to 0} \mathbb{E}_\nu[\tau_{\varepsilon,\delta}]/\log(1/\delta) \geq T_\varepsilon(\mu)$ where

$$T_\varepsilon(\mu) = \min_{i \in \mathcal{I}_\varepsilon(\mu)} \min_{\beta \in (0,1)} T_{\varepsilon,\beta}(\mu, i) \, , \quad T_{\varepsilon,\beta}(\mu,i)^{-1} = \max_{w \in \Delta_K, w_i = \beta} \min_{j \neq i} \frac{1}{2} \frac{(\mu_i - \mu_j + \varepsilon)^2}{1/\beta + 1/w_j} \, .$$

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

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

Set $\hat{\imath}_n \in \arg\max_{i \in [K]} \mu_{n,i}$, $B_n = \hat{\imath}_n$ 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 [**fixed**] $\bar{\beta}_{n+1}(i,j) = \beta$ or [**IDS**] $\beta_n(i,j) = \frac{N_{n,j}}{N_{n,i} + N_{n,j}}$;

Set $I_n = C_n$ if $N^{B_n}_{n,C_n} \leq (1 - \bar{\beta}_{n+1}(B_n, C_n)) T_{n+1}(B_n, C_n)$, otherwise set $I_n = B_n$;

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

$(N_{n,i}, \mu_{n,i})$: number of pulls and empirical mean of arm i before time n.
$T_n(i,j)$: number of selection of the leader/challenger pair (i,j) before time n.
$N^i_{n,j}$: number of pulls of arm j when selecting pair (i,j) before time n.

(ε,δ)-PAC sequential test

? How to obtain a (ε,δ)-PAC sequential test for 1-sub-Gaussian distributions ?

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

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

with $c(n,\delta) \simeq \log(1/\delta) + 2\log\log(1/\delta) + 4\log(4 + \log(n/2))$.

Asymptotic confidence guarantees

Theorem 1. *Let $\varepsilon \geq 0$ and $\varepsilon_0 > 0$. Combined with GLR$_\varepsilon$ stopping (1), the EB-TC$_{\varepsilon_0}$ algorithm satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $|i^\star(\mu)| = 1$,*

- **IDS**: $\limsup_{\delta\to 0} \mathbb{E}_\nu[\tau_{\varepsilon,\delta}]/\log(1/\delta) \leq T_{\varepsilon_0}(\mu) D_{\varepsilon,\varepsilon_0}(\mu)$,
- **fixed** $\beta \in (0,1)$: $\limsup_{\delta\to 0} \mathbb{E}_\nu[\tau_{\varepsilon,\delta}]/\log(1/\delta) \leq T_{\varepsilon_0,\beta}(\mu) D_{\varepsilon,\varepsilon_0}(\mu)$,

where $D_{\varepsilon,\varepsilon_0}(\mu) = (1 + \max_{i \neq i^\star}(\varepsilon_0 - \varepsilon)/(\mu_\star - \mu_i + \varepsilon))^2$.

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

Finite confidence guarantees

Theorem 2. *Let $\delta \in (0,1)$ and $\varepsilon_0 > 0$. Combined with GLR$_{\varepsilon_0}$ stopping (1), the EB-TC$_{\varepsilon_0}$ algorithm with fixed $\beta = 1/2$ satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ,*

$$\mathbb{E}_\nu[\tau_{\varepsilon_0,\delta}] \leq \inf_{\varepsilon \in [0,\varepsilon_0]} \max\{T_{\mu,\varepsilon_0}(\delta,\varepsilon) + 1, \, S_{\mu,\varepsilon_0}(\varepsilon)\} + 2K^2 \, , \quad \text{where}$$

$\limsup_{\delta\to 0} \frac{T_{\mu,\varepsilon_0}(\delta,0)}{\log(1/\delta)} \leq 2|i^\star(\mu)| T_{\varepsilon_0,1/2}(\mu)$, $S_{\mu,\varepsilon_0}(\frac{\varepsilon_0}{2}) = \mathcal{O}(K^2 |\mathcal{I}_{\varepsilon_0/2}(\mu)| \varepsilon_0^{-2} \log \varepsilon_0^{-1})$.

Key result: Let $\delta \in (0,1)$, $n > K$. Let $\mathcal{E}_{n,\delta}$ be a concentration event with $\mathbb{P}_\nu(\mathcal{E}_{n,\delta}^\complement) \leq K^2\delta$. Under the event $\mathcal{E}_{n,\delta}$, for all $\varepsilon \geq 0$, we have

$$\sum_{i \in \mathcal{I}_\varepsilon(\mu)} \sum_j T_n(i,j) \geq n - 8H_{\mu,\varepsilon_0}(\varepsilon)\log(n^2/\delta) - 3K^2 - 1 \, ,$$

where $H_{\mu,\varepsilon_0}(0) = \mathcal{O}(K \min\{\Delta_{\min}, \varepsilon_0\}^{-2})$ and $H_{\mu,\varepsilon_0}(\varepsilon_0/2) = \mathcal{O}(K/\varepsilon_0^2)$.

Experiments

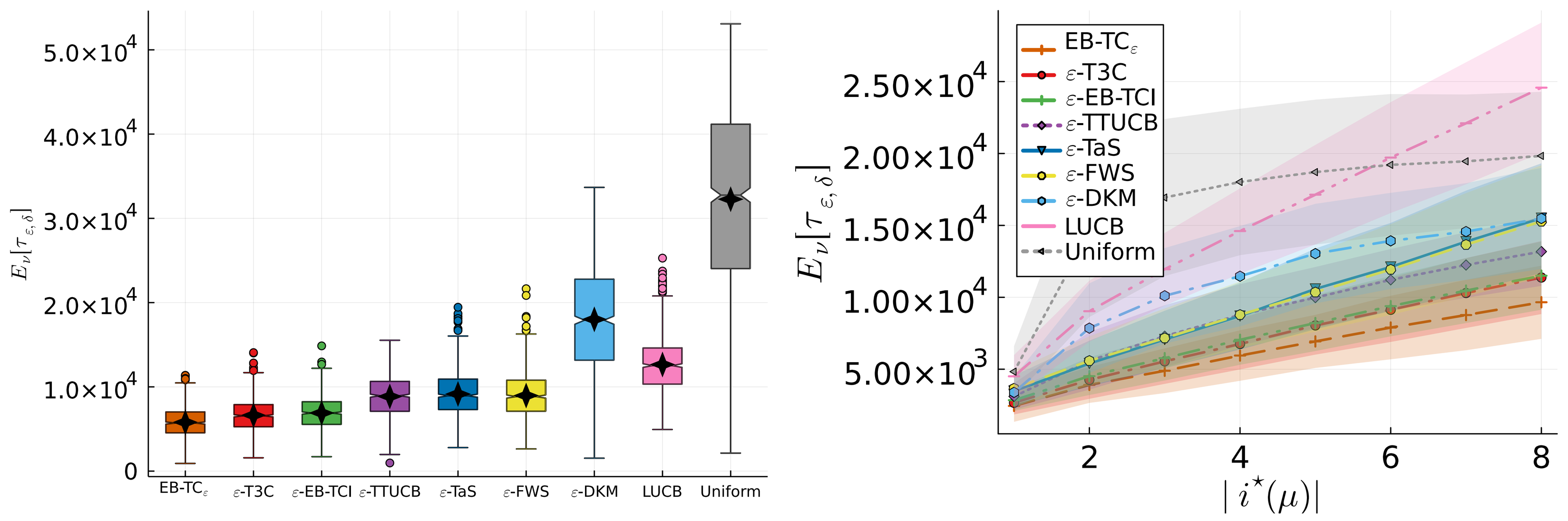

Figure 1: Stopping time on (a) random instances ($K = 20$) with $\mu_1 = 1$, $\mu_i \sim \mathcal{U}([0, 0.9])$ for all $i \geq 6$, otherwise $\mu_i \sim \mathcal{U}([0.9, 1])$ and (b) "two-groups" instances ($K = 10$) with $(\mu^\star, \Delta_2) = (0.6, 0.2)$.

Beyond fixed-confidence guarantees

Anytime guarantees on

- the **probability of ε-error** and
- the **expected simple regret**.

Theorem 3. *Let $\varepsilon_0 > 0$. The EB-TC$_{\varepsilon_0}$ algorithm with fixed proportions $\beta = 1/2$ satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ, for all $n > 5K^2/2$,*

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

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

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

Notation: distinct mean gaps $0 = \Delta_1 < \Delta_2 < \cdots < \Delta_{C_\mu} < \Delta_{C_\mu+1} = +\infty$ where $C_\mu = |\{\mu_i \mid i \in [K]\}|$. For all $\varepsilon \geq 0$, let $i_\mu(\varepsilon) = i$ if $\varepsilon \in [\Delta_i, \Delta_{i+1})$.

Other guarantees:

- **Unverifiable sample complexity**

$$\mathbb{P}_\nu(\forall n > U_{i_\mu(\varepsilon),\delta}(\mu,\varepsilon_0), \, \hat{\imath}_n \in \mathcal{I}_\varepsilon(\mu)) \geq 1 - \delta$$

where $U_{i,\delta}(\mu,\varepsilon_0) =_{\delta\to 0} 8H_i(\mu,\varepsilon_0)\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$.

- Constant **cumulative regret** in the decoupled exploration/exploitation setting.

Experiments

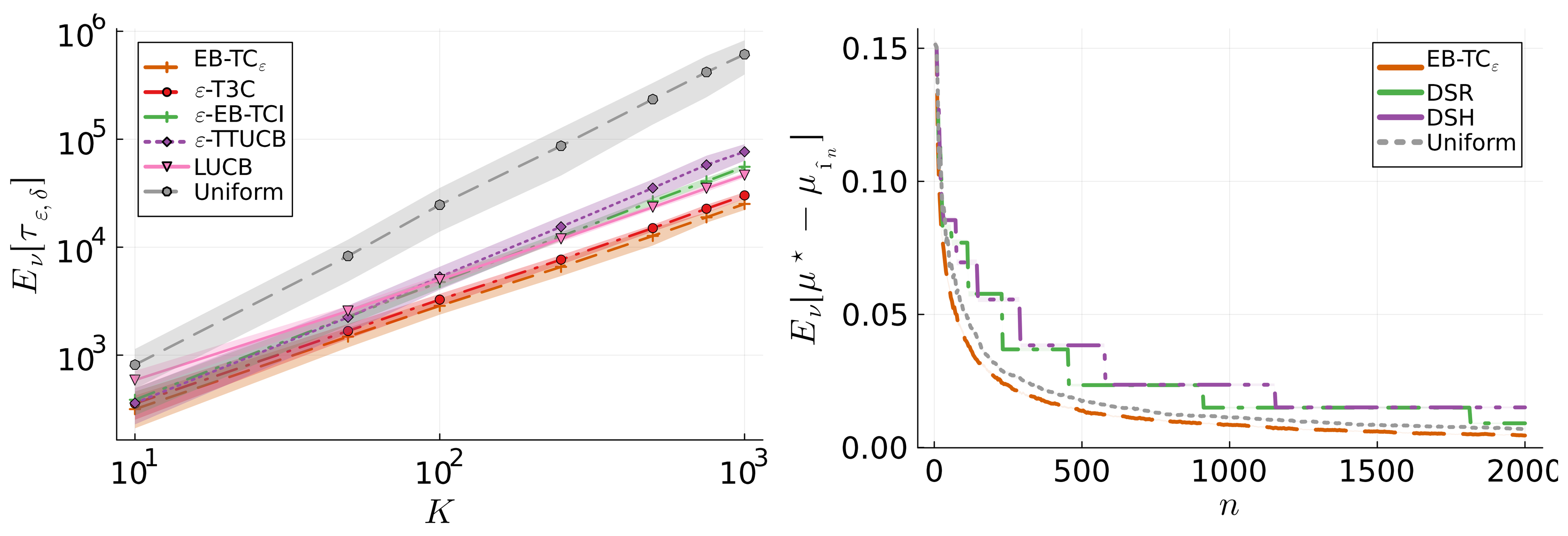

Figure 2: (a) Stopping time on instances $\mu_i = 1 - ((i-1)/(K-1))^{0.3}$ for varying K. (b) Simple regret on instance $\mu = (0.6, 0.6, 0.55, 0.45, 0.3, 0.2)$ fors EB-TC$_{\varepsilon_0}$ with slack $\varepsilon_0 = 0.1$ and fixed $\beta = 1/2$.

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

Conclusion

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. Asymptotic (β-)optimality in ε-BAI for Gaussian distributions.
4. Anytime upper bounds on the uniform ε-error and the simple regret.