Non-Asymptotic Analysis of a UCB-based Top Two Algorithm

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

Motivation

Goal: Identify the item having the highest averaged return with a given confidence.

Typical guaranty: Asymptotic optimality of the expected sample complexity.

⚠ Not informative for moderate confidence level !

☞ This paper: sample complexity upper bounds for **any confidence level** !

Best-arm identification (BAI)

K arms: arm $i \in [K]$ is associated with a Gaussian distribution $\mathcal{N}(\mu_i, 1)$.

Goal: identify $i^\star = \arg\max_{i\in[K]} \mu_i$ with confidence $1-\delta \in (0,1)$.

Algorithm: at time n,

- *Sequential test:* if the stopping time τ_δ is reached, then return the candidate answer $\hat{\imath}_n$, else
- **Sampling rule**: pull arm I_n and observe $X_n \sim \mathcal{N}(\mu_{I_n}, 1)$.

Fixed-confidence: given an confidence $\delta \in (0,1)$, define a stopping time τ_δ which is δ-correct, i.e. $\mathbb{P}_\mu(\tau_\delta < +\infty, \hat{\imath}_{\tau_\delta} \neq i^\star) \le \delta$, and

☞ Minimize the **expected sample complexity** $\mathbb{E}_\mu[\tau_\delta]$.

Lower bound on the expected sample complexity

? What is the best one could achieve ?

☞ Garivier and Kaufmann (2016): For all δ-correct algorithms and all Gaussian instances with $\mu \in \mathbb{R}^K$, $\liminf_{\delta\to 0} \mathbb{E}_\mu[\tau_\delta]/\log(1/\delta) \ge T^\star(\mu)$ where

$$T^\star(\mu) = \min_{\beta\in(0,1)} T^\star_\beta(\mu) \quad \text{and} \quad T^\star_\beta(\mu)^{-1} = \max_{w\in\triangle_K, w_{i^\star}=\beta} \min_{j\neq i^\star} \frac{1}{2}\frac{(\mu_{i^\star}-\mu_j)^2}{1/\beta + 1/w_j} .$$

TTUCB: UCB-based Top Two sampling rule

Input: fixed proportion $\beta \in (0,1)$ and function $g : \mathbb{N} \to \mathbb{R}^+$.

Get the **UCB leader** $B_n = \arg\max_{i\in[K]} \{\mu_{n,i} + \sqrt{g(n)/N_{n,i}}\}$;

Get the TC challenger $C_n \in \arg\min_{i\neq B_n} \frac{(\mu_{n,B_n} - \mu_{n,i})_+}{\sqrt{1/N_{n,B_n} + 1/N_{n,i}}}$;

Use **tracking** to get $I_n = B_n$ if $N^{B_n}_{n,B_n} \le \beta L_{n+1,B_n}$, otherwise $I_n = C_n$;

Output: next arm to sample I_n.

$(N_{n,i}, \mu_{n,i})$: number of pulls and empirical mean of arm i before time n.
$L_{n,i}$: number of selection of arm i as leader before time n.
$N^i_{n,j}$: number of pulls of arm j when arm i is leader before time n.

- Take $\beta = 1/2$ since $w^\star(\mu)_{i^\star} \le 1/2$ and $T^\star_{1/2}(\mu)/T^\star(\mu) \ll 2$ for most instances.
- Choose small g s.t. $\mathbb{P}_\mu(\mathcal{E}_n) \ge 1 - Kn^{-s}$ with

$$\mathcal{E}_n = \{\forall (t,i) \in [n^{1/\alpha}] \times [K],\ \mu_i \in [\mu_{t,i} \pm \sqrt{g(t)/N_{t,i}}]\}$$

where $\alpha, s > 1$, e.g. $g_u(n) = 2\alpha(1+s)\log n$.

δ-correct sequential test

? How to obtain a δ-correct sequential test for Gaussian distributions ?

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

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

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 $(\delta,\beta) \in (0,1)^2$. Combined with GLR stopping (1), the TTUCB algorithm is δ-correct and asymptotically β-optimal for all $\mu \in \mathbb{R}^K$ having distinct means, i.e. it satisfies $\limsup_{\delta\to 0} \mathbb{E}_\nu[\tau_\delta]/\log(1/\delta) \le T^\star_\beta(\mu)$.*

Limitations: no guarantees (1) for **moderate regime** of δ and (2) when sub-optimal arms share the **same mean**.

Finite confidence guarantees

Theorem 2. *Let $\delta \in (0,1)$. Combined with GLR stopping (1), the TTUCB algorithm using $\beta = 1/2$ and g_u with $\alpha = s = 1.2$ satisfies that, for all $\mu \in \mathbb{R}^K$ such that $|i^\star(\mu)| = 1$,*

$$\mathbb{E}_\mu[\tau_\delta] \le \inf_{x\in[0,(K-1)^{-1}]} \max\left\{T_0(\delta,x), C_\mu^{1.2}, C_0(x)^6, (2/\varepsilon)^{1.2}\right\} + 12K ,$$

where $\varepsilon \in (0,1]$ and

$$C_\mu = \mathcal{O}\left(H(\mu)\log H(\mu)\right) \quad \textit{with} \quad H(\mu) = 2\Delta_{\min}^{-2} + \sum_{i\neq i^\star} 2(\mu_{i^\star} - \mu_i)^{-2} ,$$

$$\limsup_{\delta\to 0} T_0(\delta,0)/\log(1/\delta) \le 2T^\star_{1/2}(\mu) ,$$

$$C_0(x) = 2/(\varepsilon a_\mu(x)) + 1 \quad \textit{with} \quad a_\mu(x) = (1-x)^{d_\mu(x)} \max\{\min_{i\neq i^\star} w^\star_{1/2}(\mu)_i, x/2\}$$

and $d_\mu(x) = |\{i \neq i^\star \mid w^\star_{1/2}(\mu)_i < x/2\}|$.

Refined analysis: Clipping $\min_{i\neq i^\star} w^\star_{1/2}(\mu)_i$ by $x/2$ yields $C_0(x) = \mathcal{O}(K/\varepsilon)$.

☞ Generic method that **improves the analysis of APT** (Locatelli et al, 2016).

Table 1: Upper bound on the sample complexity τ_δ in probability (§) or in expectation (†). The notation $\tilde{\mathcal{O}}$ hides polylogarithmic factors. (*) Upper bound on $\mathbb{E}_\mu[\tau_\delta \mathbb{1}(\mathcal{E})]$ where $\mathbb{P}[\mathcal{E}^{\complement}] \le \gamma$. (**) Asymptotic bound holds for instances with distinct means. Ordered references: Kalyanakrishnan et al. (2012), Karnin et al. (2013), Jamieson et al. (2014), Degenne et al. (2019), Katz-Samuels et al. (2020), Wang et al. (2021), Barrier et al. (2022).

Algorithm	Asymptotic $\delta \to 0$	Finite δ when $H(\mu) \to +\infty$
LUCB1†	$\mathcal{O}\left(H(\mu)\log(1/\delta)\right)$	$\mathcal{O}\left(H(\mu)\log H(\mu)\right)$
Exp-Gap§	$\mathcal{O}\left(H(\mu)\log(1/\delta)\right)$	$\mathcal{O}(\sum_{i\neq i^\star} \Delta_i^{-2}\log\log\Delta_i^{-1})$
lil' UCB§	$\mathcal{O}\left(H(\mu)\log(1/\delta)\right)$	$\mathcal{O}(\sum_{i\neq i^\star} \Delta_i^{-2}\log\log\Delta_i^{-1})$
DKM†	$T^\star(\mu)\log(1/\delta) + \tilde{\mathcal{O}}(\sqrt{\log(1/\delta)})$	$\tilde{\mathcal{O}}\left(KT^\star(\mu)^2\right)$
Peace§	$\mathcal{O}\left(T^\star(\mu)\log(1/\delta)\right)$	$\mathcal{O}\left(H(\mu)\log(K/\Delta_{\min})\right)$
FWS†	$T^\star(\mu)\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$	$\mathcal{O}\left(e^K H(\mu)^{19/2}\right)$
EBS†*	$T^\star(\mu)\log(1/\delta) + o(1)$	$\mathcal{O}\left(KH(\mu)^4/w_{\min}^2\right)$
TTUCB†**	$T^\star_\beta(\mu)\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$	$\mathcal{O}\left((H(\mu)\log H(\mu))^\alpha\right)$

Tracking instead of randomization

- **Fully deterministic** algorithm.
- Deterministic counts simplifies the non-asymptotic analysis.
- Faster convergence of $N_{n,i^\star}/n$ to β, at least in $\mathcal{O}(1/n)$ instead of $\mathcal{O}(1/\sqrt{n})$.

Lemma 1. *For all $n > K$ and all $i \in [K]$, we have $-1/2 \le N^i_{n,i} - \beta L_{n,i} \le 1$.*

Generic regret minimizing leader

The Top Two method is a generic wrapper to convert any regret minimization algorithm into a best arm identification strategy.

Sufficient condition: Arm $i^\star$ is leader except for a sublinear number of times.

☞ Upper bound $(N_{n,i})_{i\neq i^\star}$ or $\sum_{i\neq i^\star} \Delta_i N_{n,i}$ under a concentration event.

Lemma 2 (UCB). *Under $\mathcal{E}_n$, we have $L_{n,i^\star} \ge n - 24H(\mu)\log n - 2K - 1$.*

Experiments

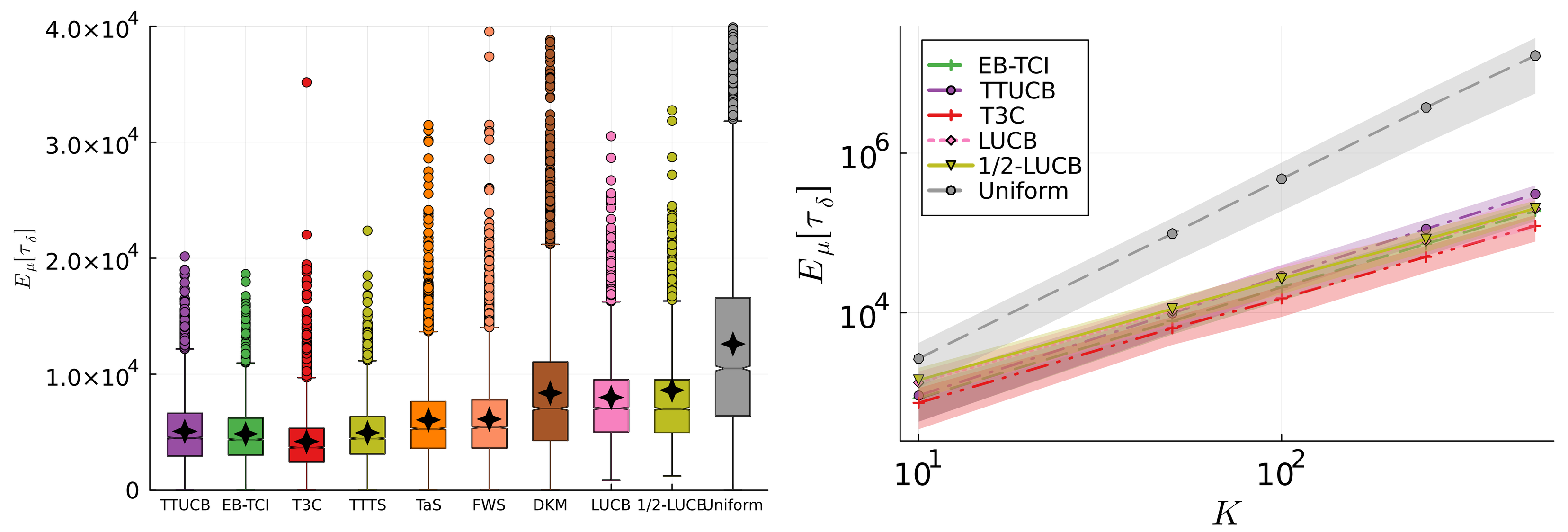

Figure 1: Empirical stopping time for $\delta = 0.1$ on (a) random instances with $\mu_1 = 0.6$ and $\mu_i \sim \mathcal{U}([0.2, 0.5])$ for $i \neq 1$ ($K = 10$) and (b) instances $\mu_i = 1 - \left(\frac{i-1}{K-1}\right)^{0.6}$ with $H(\mu) = \Theta(K^{1.2})$.

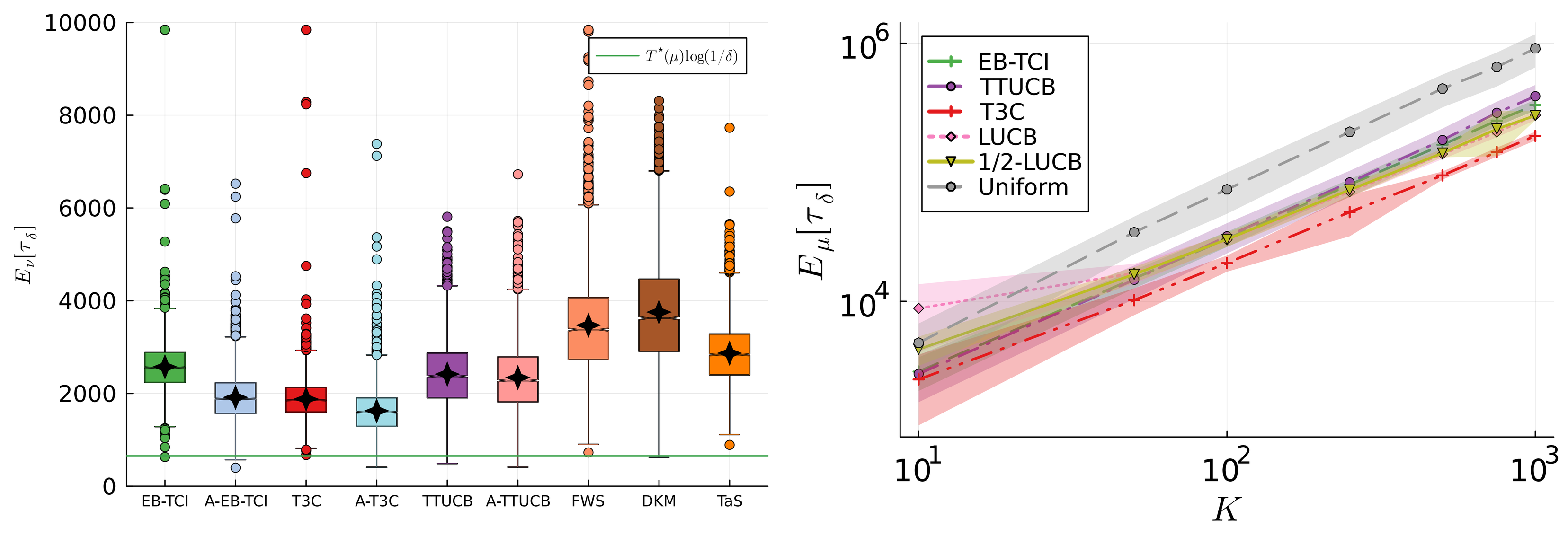

Figure 2: Empirical stopping time for $\delta = 0.1$ on "1-sparse" instances: (a) $(K, \mu_{i^\star}, \Delta) = (35, 0, 0.5)$ with $T^\star_{1/2}(\mu)/T^\star(\mu) \approx 3/2$ and (b) $(\mu_{i^\star}, \Delta) = (0, 0.25)$ with $H(\mu) = \Theta(K)$. Constant $\beta = 1/2$ and adaptive proportions (A-), IDS (You et al., 2023) sets $\beta_n = N_{n,C_n}/(N_{n,C_n} + N_{n,B_n})$.

Conclusion

1. First non-asymptotic analysis of Top Two algorithms, which holds for instances having a unique best arm.
2. Deterministic asymptotically β-optimal Top Two algorithm using UCB leader and tracking instead of randomization.