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

Marc Jourdan and Rémy Degenne

November 3, 2023

Motivation

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

Typical guaranty: Asymptotic optimality of the algorithm as regards the expected sample complexity.

⚠ Not informative for moderate confidence level !

☞ This paper:

UCB-based Top Two Algorithm with
sample complexity upper bounds for **any confidence level** !

Best-arm identification (BAI)

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

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 $X_n \sim \mathcal{N}(\mu_{I_n}, 1)$.

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

$$\mathbb{P}_\mu(\tau_\delta < +\infty, \hat{\imath}_{\tau_\delta} \neq i^\star) \leq \delta \, .$$

(Garivier and Kaufmann, 2016) For all δ-correct algorithms,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\log(1/\delta)} \geq T^\star(\mu) = \min_{\beta \in (0,1)} T^\star_\beta(\mu) \, .$$

δ-correct sequential test

? How to obtain a δ-correct sequential test for Gaussians ?

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

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

$(N_{n,i}, \mu_{n,i})$: pulling count and empirical mean of arm i before time n

TTUCB: UCB-based Top Two sampling rule

Input: proportion $\beta \in (0,1)$ and $g : \mathbb{N} \to \mathbb{R}^+$, e.g. $\beta = 1/2$ and $g_u(n) = \Theta(\log n)$.

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

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

Tracking: get $I_n = B_n$ if $N_{n,B_n}^{B_n} \leq \beta L_{n+1,B_n}$, otherwise $I_n = C_n$;

Output: next arm to sample I_n.

$L_{n,i}$: number of selection of arm i as leader before time n.
$N_{n,j}^i$: number of pulls of arm j when arm i is leader before time n.

Upper bounds on the expected sample complexity

Theorem

$\boldsymbol{\delta \to 0}$: *Combined with GLR stopping (1), TTUCB is δ-correct and asymptotically β-optimal for all instances having distinct means, i.e.*

$$\limsup_{\delta \to 0} \mathbb{E}_{\mu}[\tau_\delta] / \log(1/\delta) \le T^\star_\beta(\mu) \, .$$

$\boldsymbol{\delta \in (0,1)}$: *TTUCB with $\beta = 1/2$ and g_u satisfies that, for all instances having a* ***unique best arm****,*

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

where $\varepsilon \in (0,1]$, $\limsup_{\delta \to 0} T_0(\delta, 0)/\log(1/\delta) \le \boldsymbol{2T^\star_{1/2}(\mu)}$ *and* $\boldsymbol{C_\mu = \mathcal{O}\left(H(\mu) \log H(\mu)\right)}$ *with* $H(\mu) = 2\Delta_{\min}^{-2} + \sum_{i \neq i^\star} 2(\mu_{i^\star} - \mu_i)^{-2}$.

Empirical results

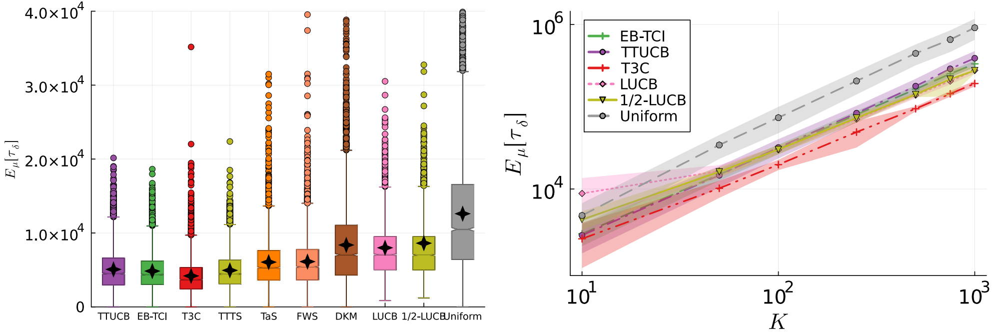

Figure: Empirical stopping time for $\delta = 0.1$ on (a) random instances with $\mu_1 = 0.6$ and $\mu_i \sim \mathcal{U}([0.2, 0.5])$ for $i \neq 1$ ($K = 10$) and (b) "1-sparse" instances $\mu_1 = 0.25$ and $\mu_i = 0$ otherwise.

Conclusion

Paper & Poster

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.