On the complexity of All ε-Best Arms Identification

Aymen Al Marjani, Tomas Kocak, Aurélien Garivier

École Normale Supérieure de Lyon, UMPA & LIP

September 20th, 2022

UNIVERSITÉ DE LYON
ENS DE LYON

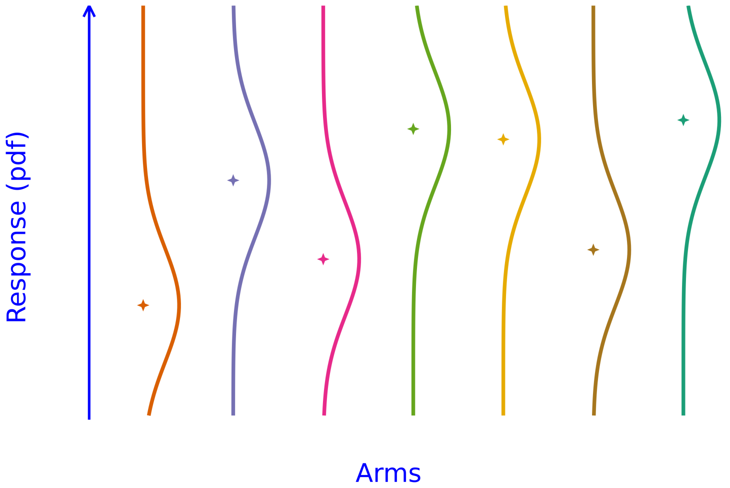
Outline

Goal: identify all ε-optimal arms

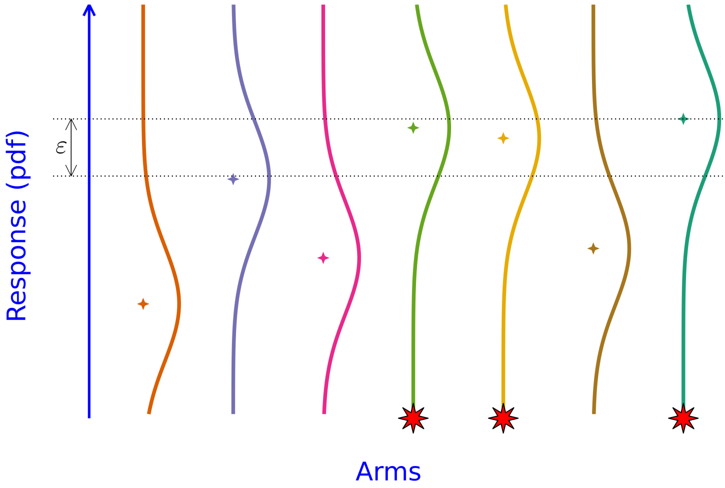
The lower bound analysis

T&S: an asymptotically optimal strategy

Multi-armed bandit model



Goal: Identify all ε-optimal arms



δ-correct Gaussian All-ε-BAI

Bandit instance: K Gaussian arms parameterized by $\boldsymbol{\mu} = (\mu_a : a \in [K])$

Sequential sampling: for $t \geq 1$, choose $A_t = \phi_t(A_1, Y_1, \dots, A_{t-1}, Y_{t-1}) \in [K]$ and observe

$$Y_t \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(\mu_{A_t}, 1)$$

Goal: for a risk $\delta \in (0, 1)$, using a number of samples τ_δ as low as possible, identify

$$G_\varepsilon(\boldsymbol{\mu}) \triangleq \left\{a \in [K] : \mu_a \geq \max_i \mu_i - \varepsilon\right\}$$

with a δ-correct algorithm outputting $\widehat{G}_\varepsilon$ depending only on the τ_δ observations obeying

$$\mathbb{P}_{\boldsymbol{\mu}}\left(\widehat{G}_\varepsilon = G_\varepsilon(\boldsymbol{\mu})\right) \geq 1 - \delta$$

Related work

- Introduced by [Mason et al., Neurips 2020]
- Example: drug selection
- $\neq$ best-arm identification and TOP-*k* arms selection
- $\neq$ ε-best-arm identification
- $\neq$ thresholding bandit

Outline

Goal: identify all ε-optimal arms

The lower bound analysis

T&S: an asymptotically optimal strategy

Complexity: Lower Bound

Theorem

For any δ-correct strategy and any bandit instance $\boldsymbol{\mu}$, the expected stopping time is lower-bounded as

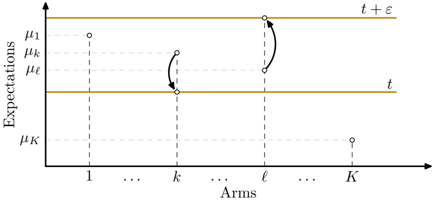
$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \geq T^*_\varepsilon(\boldsymbol{\mu}) \ \log \frac{1}{2.4\delta}$$

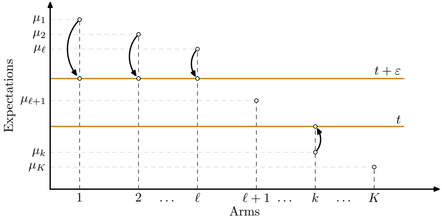
with

$$T^*_\varepsilon(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{\omega} \in \Delta_K} \underbrace{\inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{a \in [K]} \omega_a \frac{(\mu_a - \lambda_a)^2}{2}}_{T_\varepsilon(\boldsymbol{\mu},\boldsymbol{\omega})^{-1}} \qquad (\star)$$

where $\Delta_K = \left\{(\omega_1, \ldots, \omega_K) \in [0, +\infty)^K : \omega_1 + \cdots + \omega_K = 1\right\}$ is the K-simplex, and $\mathrm{Alt}(\boldsymbol{\mu})$ is the set of all bandit models with a set of ε-optimal arms different from that of $\boldsymbol{\mu}$

Solving the min problem $\boldsymbol{\lambda}^*_{\varepsilon,\boldsymbol{\mu}}(\boldsymbol{\omega}) \triangleq \arg\min_{\boldsymbol{\lambda}\in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{a\in[K]} \omega_a \frac{(\mu_a-\lambda_a)^2}{2}$





$$\boldsymbol{\lambda}^*_{\varepsilon,\boldsymbol{\mu}}(\boldsymbol{\omega}) = \arg\min_{\boldsymbol{\lambda}\in \Lambda_G\cup\Lambda_B} \sum_{a\in[K]} \omega_a \frac{(\mu_a-\lambda_a)^2}{2}$$

$$\boldsymbol{\lambda}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega}) \triangleq (\mu_1,\ldots,\underbrace{\overline{\boldsymbol{\mu}}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega})}_{\text{index } k},\ldots,\underbrace{\overline{\boldsymbol{\mu}}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega})+\varepsilon}_{\text{index } \ell},\ldots,\mu_K)^\mathsf{T} \text{ for } k\in G(\boldsymbol{\mu})$$

$$\boldsymbol{\lambda}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega}) \triangleq (\underbrace{\overline{\boldsymbol{\mu}}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega})+\varepsilon}_{\text{indices } 1 \text{ to } \ell},\mu_{\ell+1},\ldots,\underbrace{\overline{\boldsymbol{\mu}}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega})}_{\text{index } k},\ldots,\mu_K)^\mathsf{T} \text{ for } k\notin G(\boldsymbol{\mu}) \qquad \boldsymbol{\mu}^{k,\ell}_{\epsilon}(\boldsymbol{\omega}) = \frac{\omega_k\mu_k+\omega_\ell(\mu_\ell-\varepsilon)}{\omega_k+\omega_\ell}$$

$$\Lambda_G = \left\{\boldsymbol{\lambda}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega}) : k\in G_\varepsilon(\boldsymbol{\mu}), \ell\in G_\varepsilon(\boldsymbol{\mu})\setminus\{k\}\right\},$$

$$\Lambda_B = \left\{\boldsymbol{\lambda}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega}) : k\notin G_\epsilon(\boldsymbol{\mu}), \ell\in [|1,k-1|] \quad \text{s.t. } \mu_\ell \geq \boldsymbol{\mu}^{k,\ell}_{\varepsilon}(\boldsymbol{\omega})+\varepsilon > \mu_{\ell+1}\right\}$$

Computing the optimal weights

$$T_\varepsilon(\boldsymbol{\mu}, \boldsymbol{\omega})^{-1} = \inf_{\boldsymbol{d} \in \mathcal{D}_{\varepsilon, \boldsymbol{\mu}}} \boldsymbol{\omega}^\intercal \boldsymbol{d} \qquad (1)$$

where

$$\mathcal{D}_{\varepsilon, \boldsymbol{\mu}} \triangleq \left\{ \left(\frac{(\lambda_a - \mu_a)^2}{2} \right)_{a \in [K]}^\intercal \quad : \quad \boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu}) \right\}$$

Danskin's theorem: let $\boldsymbol{\lambda}^*(\boldsymbol{\omega})$ be a best response to $\boldsymbol{\omega}$ and define $\boldsymbol{d}^*(\boldsymbol{\omega}) \triangleq \left(\frac{(\boldsymbol{\lambda}^*(\boldsymbol{\omega})_a - \mu_a)^2}{2}\right)_{a \in [K]}^\intercal$, then $\boldsymbol{d}^*(\boldsymbol{\omega})$ is a supergradient of $T_\varepsilon(\boldsymbol{\mu}, .)^{-1}$ at $\boldsymbol{\omega}$

Besides, the function $\boldsymbol{\omega} \mapsto T_\varepsilon(\boldsymbol{\mu}, \boldsymbol{\omega})^{-1}$ is L-Lipschitz with respect to $\| \cdot \|_1$ for

$$L \geq \max_{a,b \in [K]} \frac{(\mu_a - \mu_b + \varepsilon)^2}{2}$$

Mirror ascent

For a (convex) miror map Φ and a learning rate $(\alpha_n)_n$, mirror ascent is defined as:

$$\omega_{n+1} = \nabla\Phi^{-1}\Big(\nabla\Phi(\omega_n) + \alpha_n \nabla f(\omega_n)\Big)$$

Theorem [e.g. Bubeck '2015]

Let $\omega_1 = (\frac{1}{K}, \ldots, \frac{1}{K})^\top$ and learning rate $\alpha_n = \frac{1}{L}\sqrt{\frac{2\log K}{n}}$. The mirror ascent algorithm defined on the simplex Δ_K with as a mirror map the generalized negative entropy $\Phi(\omega) = \sum_{a\in[K]} \omega_a \log(\omega_a)$ enjoys the following guarantees:

$$f(\omega^*) - f\left(\frac{1}{N}\sum_{n=1}^{N}\omega_n\right) \leq \frac{2}{\max_{a,b\in[K]}(\mu_a - \mu_b + \varepsilon)^2}\sqrt{\frac{2\log K}{N}}$$

About the moderate confidence regime

$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \geq T^*_\varepsilon(\boldsymbol{\mu}) \ \log \frac{1}{2.4\delta}$ is tight when $\delta \to$, what about $\delta \approx 1/10$?

Theorem

Fix $\delta \leq 1/10$ and $\varepsilon > 0$. Consider an instance ν such that there exists at least one bad arm: $G_\varepsilon(\boldsymbol{\mu}) \neq [K]$. Wlog, suppose that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$ and define the lower margin $\beta_\varepsilon = \min_{k \notin G_\varepsilon(\boldsymbol{\mu})} \mu_1 - \varepsilon - \mu_k$.

Then any δ-PAC algorithm has an average sample complexity over all permuted instances satisfying

$$\mathbb{E}_{\pi \sim \mathbf{S}_K} \mathbb{E}_{\pi(\boldsymbol{\mu})}[\tau_\delta] \geq \frac{1}{12|G_{\beta_\varepsilon}(\boldsymbol{\mu})|^3} \sum_{b=1}^{K} \frac{1}{(\mu_1 - \mu_b + \beta_\varepsilon)^2},$$

$\to \tau_\delta$ is linear in K (higher bound in some settings)

UNIVERSITÉ DE LYON

ENS DE LYON

Outline

Goal: identify all ε-optimal arms

The lower bound analysis

T&S: an asymptotically optimal strategy

Sampling rule

Denoting $\boldsymbol{N}_a(t) = \sum_{s \leq t} \mathbb{1}\{A_s = a\}$ the current number of draws, estimate of the means:

$$\widehat{\boldsymbol{\mu}}_t = \boldsymbol{N}_a(t)^{-1} \sum_{s : A_s = a} Y_s$$

$\rightarrow$ ($1/\sqrt{t}$-approximate) estimate of the optimal frequencies

$$\widetilde{\boldsymbol{\omega}}(\widehat{\boldsymbol{\mu}}_t) \text{ s.t. } T^*_\varepsilon(\hat{\boldsymbol{\mu}}_t) = T_\varepsilon(\hat{\boldsymbol{\mu}}_t, \widetilde{\boldsymbol{\omega}}(\widehat{\boldsymbol{\mu}}_t))$$

$\widetilde{\boldsymbol{\omega}}^{\eta_t}(\widehat{\boldsymbol{\mu}}_t)$ = projection onto $\Delta_K \cap [\eta_t, 1]^K$ for $\eta_t^{-1} = 2\sqrt{K^2 + t}$ (forced exploration)

Track the optimal proportions:

$$A_{t+1} = \arg\min_a \boldsymbol{N}_a(t) - \sum_{s=1}^{t} \widetilde{\boldsymbol{\omega}}_a^{\eta_t}(\widehat{\boldsymbol{\mu}}_s)$$

Prop: $\boldsymbol{N}_a(t) \sim \omega_a^*(\boldsymbol{\mu})t$ for all $a \in [K]$ when $t \to \infty$.

Stopping rule

Generalized Likelihood Ratio test: the statistic can be written

$$Z(t) = t \times T_\varepsilon\left(\widehat{\boldsymbol{\mu}}_t, \frac{\boldsymbol{N}(t)}{t}\right)^{-1}$$

where $\boldsymbol{N}(t) = \big(N_a(t)\big)_{a\in[K]}$

Stopping time

$$\tau_\delta = \inf\left\{t \in \mathbb{N} \ : \ Z(t) > \beta(t,\delta)\right\}$$

$\beta(\delta, t) \approx \log(1/\delta) + \frac{K}{2}\log(\log(t/\delta))$ is enough to ensure that

$$\mathbb{P}_{\boldsymbol{\mu}}\left(G_\varepsilon(\widehat{\boldsymbol{\mu}}_{\tau_\delta}) \neq G_\varepsilon(\boldsymbol{\mu})\right) \leq \delta$$

Asymptotic optimality of Track-and-Stop

Theorem (See [Garivier&Kaufmann, COLT'2016])

For all $\delta \in (0,1)$, Track-and-Stop terminates almost-surely and its stopping time τ_δ satisfies:

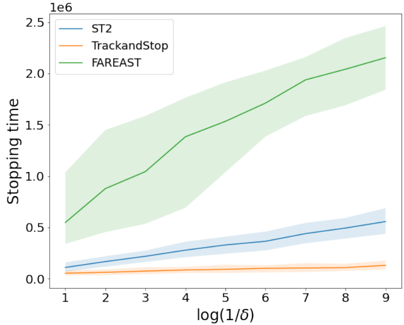
$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leq T_\varepsilon^*(\boldsymbol{\mu})^{-1}$$

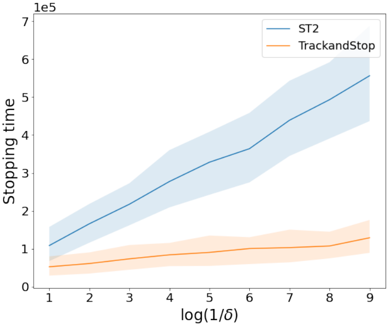
$\implies$ T&S matches the lower bound for small δ

in practice, very good even for moderate δ unless $K \gg 1$ (see below)

For non-asymptotic bounds (and algorithms), see [Barrier et al., AISTATS'22]

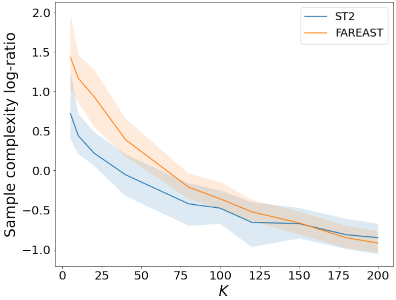
Experiment 1: small δ





$\boldsymbol{\mu} = [1, 1, 1, 1, 0.05]$, $\varepsilon = 0.9$, $N = 100$ Monte-Carlo simulations for each risk level, 10% and 90% quantiles (shaded area) for each algorithm.
Comparison with FAREAST and $(\mathrm{ST})^2$ from [Mason et al, 2020]

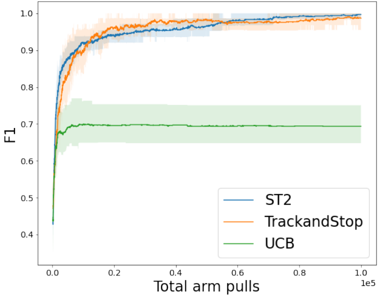
Experiment 2: moderate confidence



$\forall a \in [|1, K-1|],\ \mu_a = 1$ and $\mu_K = 0.05$.
$\varepsilon = 0.9$, $\delta = 0.1$, $N = 30$ Monte-Carlo simulations for each K

$\rightarrow$ above $\approx K = 50$ arms, the complexity is driven by the moderate regime for which FAREAST and $(\mathrm{ST})^2$ are better suited

Experiment 3: Cancer Drug Discovery experiment [Mason et al, 2020]



Goal = find among a list of 189 chemical compounds potential inhibitors to **ACRVL1**, a kinase that has been linked to several forms of cancer.

Fixed budget $N = 10^5$, mutiplicative $\varepsilon = 0.8$.

F1 score = harmonic mean of precision and recall

$\rightarrow (\mathrm{ST})^2$ and Track-and-Stop have comparable performance and that both outperform UCB's sampling scheme.

UNIVERSITÉ DE LYON

ENS DE LYON

Conclusion

$\Longrightarrow$ New sample complexity analysis of all-epsilon BAI

$\Longrightarrow$ Optimal lower bound in the asymptotic regime $\delta \to 0$

$\Longrightarrow$ sub-optimal bound for the moderate regime case that is relevant in particular when *K* $\gg 1$

$\Longrightarrow$ Computationnally efficient Track-and-Stop strategy

$\Longrightarrow$ Theoretical and practical improvement over FAREAST and $(\mathrm{ST})^2$ algorithms (= state-of-the-art for this problem).

$\Longrightarrow$ Optimality in the moderate confidence regime remains to be understood

UNIVERSITÉ DE LYON

ENS DE LYON