Top-m identification for linear bandits

Clémence Réda[1,*], Émilie Kaufmann[2], Andrée Delahaye-Duriez[1,3,4]

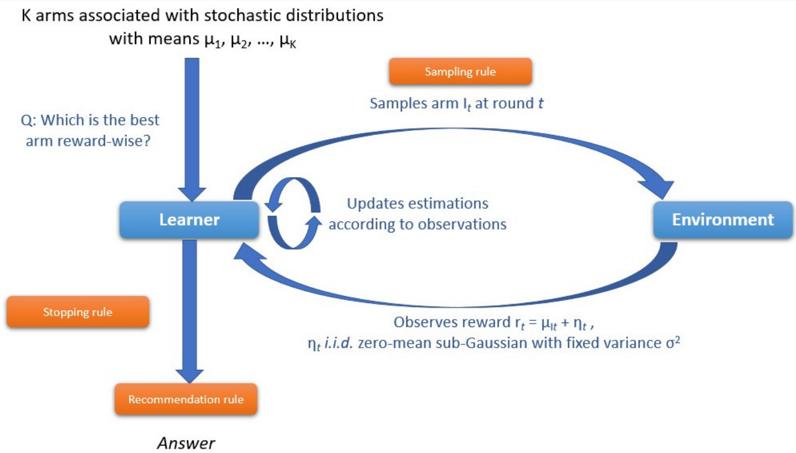
AISTATS 2021

[1] Université de Paris, Inserm UMR 1141 NeuroDiderot, Paris, France
[2] Université Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRIStAL, Lille, France
[3] Université Sorbonne Paris Nord, UFR SMBH, Bobigny, France
[4] AP-HP, Hôpital Jean Verdier, Service d'Histologie-Embryologie-Cytogénétique, Bondy, France

[*] clemence.reda@inria.fr

Pure exploration in a multi-armed bandit (MAB) setting

Université de Paris · Inserm · NEURODIDEROT · cnrs · CRIStAL



Fixed-confidence Top-m identification in a linear setting

Arms $1, 2, \dots, K < +\infty$ with features $x_1, x_2, \dots, x_K \in \mathbb{R}^N$

means $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m > \mu_{m+1} \geq \cdots \geq \mu_K \in \mathbb{R}$

(ε, m, δ)-PAC algorithms, $\delta \in (0,1), \varepsilon \geq 0$

Returns $\hat{\mathcal{S}}_m^t \subseteq \{1, \dots, K\} = [K]$ of size m at round t such that

$$\mathbb{P}(\hat{\mathcal{S}}_m^t \subseteq \mathcal{S}_m^{\star\varepsilon}) \geq 1 - \delta \text{ (error less than } \delta)$$

where $\mathcal{S}_m^{\star\varepsilon} \triangleq \{a \in [K] : \mu_a \geq \overset{m}{\max_{b \in [K]}} \mu_b - \varepsilon\}$ and $\exists \theta \in \mathbb{R}^N, \forall a \in [K], \mu_a = \theta^\top x_a$

- Prior work in a featureless setting (LUCB[1], UGapE[2])
- Prior work in a linear setting for **m=1**[3]

[1] Kalyanakrishnan, S., Tewari, A., Auer, P., & Stone, P. (2012). *ICML*.

[2] Gabillon, V., Ghavamzadeh, M., & Lazaric, A. (2012). *NIPS*.

[3] Soare, M., Lazaric, A., & Munos, R. (2014). *NIPS*; Xu, L., Honda, J., & Sugiyama, M. (2018). *AISTATS*.

GIFA (Gap-Index Focused Algorithms) family for Top-m

```
t ← 1
while stopping(t) > ε do
    // estimated m best arms at t
    J(t) ← compute_Jt(t)
    // estimated m-best arm at t
    b_t ← compute_bt(t)
    // challenger to b_t at t
    c_t ← arg max^1_{a∉J(t)} B_{a,b_t}(t)
    // sampling arms
    I_t ← selection(b_t, c_t)
    r_t ← sample(I_t)
    Update gap indices B_{i,j}(t+1)
    t ← t + 1
end while
Ŝ^t_m ← J(t)
return Ŝ^t_m
```

Rely on *good gap indices* $B_{i,j}(t)$

$$\mathbb{P}\left(\mathcal{E}_m\right) \geq 1 - \delta$$

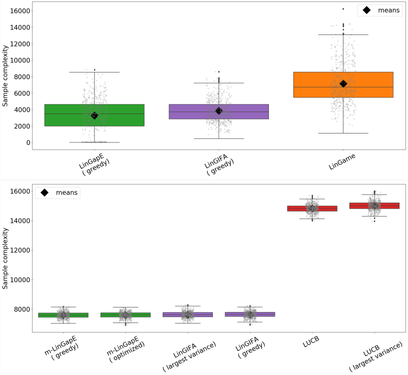
$$\mathcal{E}_m \triangleq \bigcap_{t>0} \bigcap_{j \in (\mathcal{S}_m^{\star\varepsilon})^c} \bigcap_{k \in \mathcal{S}_m^{\star 0}} \left(B_{k,j}(t) \geq \mu_k - \mu_j\right)$$

Core idea:

1. estimate the best set $J(t)$ at each time t
2. refine it by comparing two ambiguous arms $b_t \in J(t)$ and $c_t \notin J(t)$

Comparison to the state-of-the-art

Université de Paris · Inserm · NEURODIDEROT · CNRS · CRIStAL



Still performant in $m = 1$ instance w.r.t. LinGame[a]

In drug repurposing instance ($m = 5$), improvement by a factor $\frac{1}{2}$ w.r.t. LUCB[b]

[a] Degenne, R., Ménard, P., Shang, X., & Valko, M. (2020). *ICML*.

[b] Kalyanakrishnan, S., Tewari, A., Auer, P., & Stone, P. (2012). *ICML*.

References i

Degenne, R., Ménard, P., Shang, X., & Valko, M. (2020). Gamification of pure exploration for linear bandits. *ICML*, *119*, 2432–2442.

Gabillon, V., Ghavamzadeh, M., & Lazaric, A. (2012). Best arm identification: A unified approach to fixed budget and fixed confidence. *NIPS*, *25*, 3212–3220.

Kalyanakrishnan, S., Tewari, A., Auer, P., & Stone, P. (2012). Pac subset selection in stochastic multi-armed bandits. *ICML*, 227–234.

Soare, M., Lazaric, A., & Munos, R. (2014). Best-arm identification in linear bandits. *NIPS*, *27*, 828–836.

References ii

Xu, L., Honda, J., & Sugiyama, M. (2018). A fully adaptive algorithm for pure exploration in linear bandits. *AISTATS*, *84*, 843–851.