Best-Arm Identification in Unimodal Bandits

Riccardo Poiani[1], Marc Jourdan[2], Rémy Degenne[3], Emilie Kaufmann[3]
[1] Bocconi University, Milan, Italy, [2] EPFL, Lausanne, Switzerland, [3] Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189-CRIStAL

Unimodal Bandits

Problem Setting

- **Input:** Arms $(\nu_i)_{i\in[K]}$ of distributions with means μ_i
- **Unimodality:** $\exists\star \in [K]$ such that $\mu_i \leq \mu_{i+1}$ for all $i < \star$ and $\mu_i > \mu_{i+1}$ for all $i \geq \star$
- **Goal:** Identify $\star \in \operatorname{argmax}_{i\in[K]} \mu_i$ with probability at least $1-\delta$ while minimizing the *sample complexity*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] = \sum_{i\in[K]} \mathbb{E}_{\boldsymbol{\mu}}[N_i(\tau_\delta)]$$

Dependency in the Number of Arms

(Another) Lower Bound

Thm. Let $\Delta > 0$, and $\boldsymbol{\nu}^{(i)} = \mathcal{N}(\boldsymbol{\mu}^{(i)}, I_K)$ where $\mu_i^{(i)} = \Delta$ and $\mu_j^{(i)} = 0$ if $j \neq i$. For $\delta \leq 0.25$, we have:

$$\frac{1}{K}\sum_{i\in[K]} \mathbb{E}_{\boldsymbol{\nu}^{(i)}}[\tau_\delta] \geq \frac{K}{64\Delta^2}.$$

→ $\exists \boldsymbol{\nu}^{(i)}$ such that $\mathbb{E}_{\boldsymbol{\nu}^{(i)}}[\tau_\delta] \geq \frac{K}{64\Delta^2}$

👎 In the **worst-case**, there is a linear dependency in K

Unimodal TaS

How to apply Track and Stop? [2]

- We can compute $\boldsymbol{\omega}^*$ only for unimodal $\boldsymbol{\mu}$'s
- **Project** $\hat{\boldsymbol{\mu}}(t) \to \tilde{\boldsymbol{\mu}}(t)$ to be unimodal!

👍 Projection error to 0 → Optimality
👍 Projection takes $\mathcal{O}(K)$
👎 Forced Exploration **do not exploit sparsity**!

Experiments Summary

- All our algorithms outperform asymptotic optimal algorithms for generic structures [1, 4, 5]
- U-TaS suffers a lot when K is large and $\boldsymbol{\mu}$ is not flat

References

[1] Rémy Degenne, Wouter M Koolen, and Pierre Ménard. Non-asymptotic pure exploration by solving games. *Advances in Neural Information Processing Systems*, 32, 2019.

[2] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR, 2016.

[3] Marc Jourdan and Rémy Degenne. Non-asymptotic analysis of a ucb-based top two algorithm. *Advances in Neural Information Processing Systems*, 36:68980–69020, 2023.

[4] Pierre Ménard. Gradient ascent for active exploration in bandit problems. *arXiv preprint arXiv:1905.08165*, 2019.

[5] Po-An Wang, Ruo-Chun Tzeng, and Alexandre Proutiere. Fast pure exploration via frank-wolfe. *Advances in Neural Information Processing Systems*, 34:5810–5821, 2021.

Dependency in the Risk Parameter δ

(Instance-Dependent) Sample Complexity Lower Bound

Thm. For any δ-correct strategy, and any unimodal bandit $\boldsymbol{\mu}$ it holds that $\mathbb{E}_{\boldsymbol{\mu}}[c_{\tau_\delta}] \geq T^*(\boldsymbol{\mu}) \log\left(\frac{1}{2.4\delta}\right)$ where

$$T^*(\boldsymbol{\mu})^{-1} := \sup_{\boldsymbol{\omega}\in\tilde{\Delta}_K(\boldsymbol{\mu})} \min_{i\in\mathcal{N}(\star)} g_i(\boldsymbol{\omega},\boldsymbol{\mu}), \qquad g_i(\boldsymbol{\omega},\boldsymbol{\mu}) = \inf_{x\in(\mu_i,\mu_\star)} \omega_\star d(\mu_\star, x) + \omega_i d(\mu_i, x)$$

$$\tilde{\Delta}_K(\boldsymbol{\mu}) = \{\boldsymbol{\omega}\in\Delta_K | \forall i \notin \mathcal{N}(\star)\cup\star, \omega_i = 0\}$$

and $\mathcal{N}(\star)$ are the neighbors of arm $\star$.

Remarks

- Sparsity pattern of the oracle-weights! No dependency in $T^*(\boldsymbol{\mu})^{-1}$ on K.
- Lower bound is exactly BAI lower bound but on arms in $\mathcal{N}(\star)\cup\star$, i.e., [2]
- We have fast algorithms for computing $\boldsymbol{\omega}^*(\boldsymbol{\mu})$ (e.g., bisection methods)

Stopping Rules

Full-sum Stopping Rule

$$\inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(\hat{\boldsymbol{\mu}}(t))} \sum_{i\in[K]} N_i(t) d(\hat{\mu}_i(t), \lambda_i) \geq c_K(t-1,\delta), \qquad c_K(t,\delta) \approx \log\frac{1}{\delta} + K\log t$$

- The l.h.s. can be very large, although only 3 arms matter
- Cannot use the empirical threshold $\tilde{c}(t,\delta) = \log\left(\frac{1+\log t}{\delta}\right)$ commonly used in experiments
- Gaussian bandits with variance σ^2. Arm means 0 but one with mean Δ. After init, l.h.s. $\approx K D_\sigma$

Local GLR Stopping Rule

$$\imath_t = \operatorname*{argmax}_{k\in[K]} \min_{j\in\mathcal{N}(i)} W_t(i,j), \qquad W_t(i,j) = \inf_{\lambda_j > \lambda_i} \sum_{k\in\{i,j\}} N_k(t) d(\hat{\mu}_k(t), \lambda_k)$$

$$\min_{j\in\mathcal{N}(\imath_t)} W_t(\imath_t, j) \geq c(t-1,\delta), \qquad c(t-1,\delta) \approx \log\left(\frac{K}{\delta}\right) + \log(t)$$

- We exploit **local** information to stop
- No linear dependency in K, but logarithmic (due to union bound)

Optimistic Track and Stop

How to apply Optimistic Track and Stop? [1]
✖ **Confidence Intervals**

$$\Theta_t := \{\boldsymbol{\theta} | \forall i\in[K] : N_i(t) d(\hat{\mu}_i(t), \theta_i) \leq f(t)\}, \qquad f(t) \approx \log(t)$$

✔ **Structured Confidence Intervals** $\tilde{\Theta}_t = \Theta_t \cap \mathcal{S}$ where $\mathcal{S}$ are unimodal bandits

</> Compute

$$\boldsymbol{\mu}^+(t), \boldsymbol{\omega}(t) \in \operatorname*{argmax}_{\boldsymbol{\lambda}\in\tilde{\Theta}_t} \max_{\boldsymbol{\omega}\in\tilde{\Delta}_K(\boldsymbol{\lambda})} \min_{i\in\mathcal{N}(i^*(\boldsymbol{\lambda}))} g_i(\boldsymbol{\omega},\boldsymbol{\lambda})$$

👍 Can be computed in $\mathcal{O}(K)$ using K calls to $T^*(\cdot)^{-1}$
👍 There exists a time on the good event where O-TaS pulls only arms within $\star\cup\mathcal{N}(\star)$
👎 Asymptotically optimal but finite time bound does not match dependency in K

Unimodal Top-Two Sampling

How to apply Top-Two Approaches? [3]
✔ **(Structured) Leader**

$$B_t = \operatorname*{argmax}_{i\in[K]} \max_{\boldsymbol{\lambda}\in\tilde{\Theta}_t} \lambda_i$$

✔ **(Unimodal) Challenger**

$$C_t = \operatorname*{argmin}_{j\in\mathcal{N}(B_t)} W_t(B_t, j)$$

</> Arms are pulled according to fixed-design

👍 Computing the leader takes $\mathcal{O}(K)$ and no calls to $T^*(\cdot)^{-1}$
👍 There exists a time on the good event after which the leader is always $\star$
👎 Asymptotically β-optimal with fixed design, finite time bound is $\widetilde{O}(K)$