On Best-Arm Identification with a Fixed Budget in Non-Parametric Multi-Armed Bandits

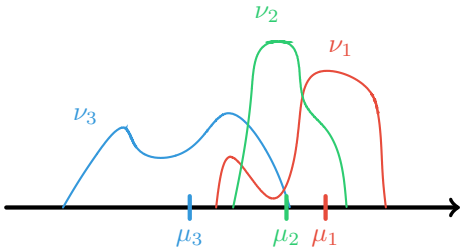
Antoine Barrier[1,2] · Aurélien Garivier[1,3] · Gilles Stoltz[2]

1. UMPA, ÉNS de Lyon 2. LMO, Université Paris-Saclay 3. LIP, ÉNS de Lyon

34th International Conference on Algorithmic Learning Theory
February 21, 2023 · Singapore

Bandit algorithms

- Model $\mathcal{D}$: set of distributions
  (*e.g.* Gaussian or Bernoulli distributions, exponential model, $\mathcal{P}[0,1]$, ...)
- K arms: K *unknown* distributions $\underline{\nu} = (\nu_1, \dots, \nu_K)$ of $\mathcal{D}$, $\mathrm{E}(\nu_a) = \mu_a$



- At round t:
  - choose an arm A_t depending on previous observations
  - observe a new independent reward $Y_t \sim \nu_{A_t}$

$$a^\star(\underline{\nu}) = \mathrm{argmax}_{1 \le a \le K} \mu_a \qquad \nu^\star = \nu_{a^\star(\underline{\nu})} \qquad \mu^\star = \mu_{a^\star(\underline{\nu})}$$

Regret minimization: optimal non-parametric approach

$$\textbf{Minimize} \quad R_T(\underline{\nu}) = \sum_{t=1}^{T} \mu^\star - \mathbb{E}_{\underline{\nu}}[\mu_{A_t}] = \sum_{\substack{1 \le a \le K \\ a \neq a^\star(\underline{\nu})}} \mathbb{E}_{\underline{\nu}}[N_a(T)]\,\Delta_a$$

with $N_a(T)$: # of pulls of arm a at time T ; $\Delta_a = \mu^\star - \mu_a$: *gap* of arm a

$\rightarrow$ **Non-parametric inequalities based on** $\mathcal{K}_{\inf}$: for $\nu \in \mathcal{D}$ and $x \in \mathbb{R}$

$$\mathcal{K}_{\inf}^{>}(\nu, x) = \inf\big\{\mathrm{KL}(\nu, \zeta) : \zeta \in \mathcal{D} \ \text{ s.t. } \ \mathrm{E}(\zeta) > x\big\}$$

Lower bound [Lai and Robbins, 1985, Burnetas and Katehakis, 1996]

$$\forall a \neq a^\star(\underline{\nu}), \qquad \liminf_{T \to +\infty} \frac{\mathbb{E}_{\underline{\nu}}[N_a(T)]}{\ln T} \geq \frac{1}{\mathcal{K}_{\inf}^{>}(\nu_a, \mu^\star)}$$

Upper bounds Gap-based (*e.g.* UCB) and then $\mathcal{K}_{\inf}^{>}$-based (*e.g.* KL-UCB-switch) guarantees

Best-Arm Identification (BAI) with a fixed confidence

Return an estimate $\hat{a}_\tau$ **of** $a^\star(\underline{\nu})$ **after some stopping time** τ

$$\textbf{Minimize } \mathbb{E}_{\underline{\nu}}[\tau] \qquad \textbf{such that} \quad \mathbb{P}_{\underline{\nu}}\big(\hat{a}_\tau \neq a^\star(\underline{\nu})\big) \leq \delta$$

Lower bound [Garivier and Kaufmann, 2016, Jourdan et al., 2022]

$$\mathbb{E}_{\underline{\nu}}[\tau] \geq T(\underline{\nu}) \ln(1/\delta)$$

$$T(\underline{\nu})^{-1} = \sup_{w \in \Sigma_K} \min_{a \neq a^\star(\underline{\nu})} \inf_{x \in [\mu_a, \mu^\star]} w_{a^\star(\underline{\nu})}\, \mathcal{K}_{\inf}^{<}(\nu^\star, x) + w_a\, \mathcal{K}_{\inf}^{>}(\nu_a, x)$$

Upper bound Asymptotically ($\delta \to 0$)

- **optimal strategies** for exponential models
- **almost optimal strategies** for non-parametric models

$$\rightarrow \textbf{ Non-parametric theory based on } \mathcal{K}_{\inf}$$

Best-Arm Identification (BAI) with a fixed budget

Gap-based approaches [Audibert et al., 2010]

Minimize $\mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big)$ **given budget** T

Order arms by their means: $\mu_{(1)} > \mu_{(2)} > \mu_{(3)} > \cdots > \mu_{(K)}$

Lower bound For the model $\mathcal{D} = \mathcal{B}_{[p,1-p]} = \big\{\mathrm{Ber}(x) : x \in [p, 1-p]\big\}$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \geq -\frac{5}{p(1-p)} \min_{2\leq k\leq K} \frac{\Delta_{(k)}^2}{k}$$

Upper bound For the model $\mathcal{D} = \mathcal{P}[0,1]$ (or subGaussian): the successive-rejects strategy satisfies:

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \leq -\frac{1}{\overline{\ln}\, K} \min_{2\leq k\leq K} \frac{\Delta_{(k)}^2}{k}$$

See also [Karnin et al., 2013, Kaufmann et al., 2016, Carpentier and Locatelli, 2016]

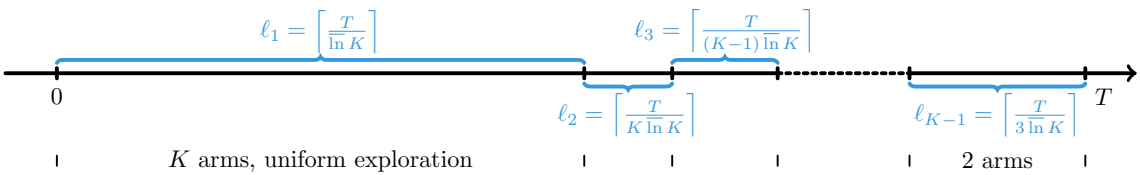
→ **Gap-based** theory for **specific models**

→ **What about non-parametric bounds?**

The successive-rejects strategy

- Exploration is split in $K-1$ **phases** of various lengths $\ell_1, \ldots, \ell_{K-1}$.
- At each phase, all remaining arms are pulled uniformly and the arm with worst empirical mean is removed.

Phase lengths of [Audibert et al., 2010]:



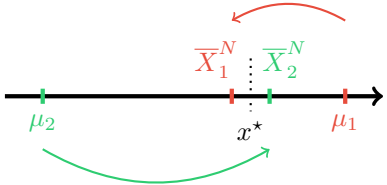
$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \leq -\frac{1}{\overline{\ln} K} \min_{2\leq k\leq K} \frac{\Delta_{(k)}^2}{k}$$

$\rightarrow$ **Proof based on Hoeffding's inequality**
to bound errors when comparing empirical means

Non-parametric upper bound

$$\mathcal{D} = \mathcal{P}[0,1]$$

$$\Delta = \mu_1 - \mu_2 > 0$$



~~**Hoeffding's inequality**~~: $\frac{1}{N}\log\mathbb{P}\left(\overline{X}_2^N \geq \overline{X}_1^N\right) \leq -\Delta^2$

Cramér-Chernoff bound separately for $\overline{X}_1^N$ and $\overline{X}_2^N$ with the **best cut-off** $x^\star$:

$$\limsup_{N\to+\infty} \frac{1}{N}\log\mathbb{P}\left(\overline{X}_2^N \geq \overline{X}_1^N\right) \leq -\left(\mathcal{L}_{\inf}^{>}(x^\star,\nu_2) + \mathcal{L}_{\inf}^{<}(x^\star,\nu_1)\right) \stackrel{\text{def}}{=} -\mathcal{L}(\nu_2,\nu_1)$$

where, for $\nu \in \mathcal{D}$ and $x \in \mathbb{R}$:

$$\mathcal{L}_{\inf}^{<}(x,\nu) = \inf\left\{\mathsf{KL}(\zeta,\nu) : \zeta \in \mathcal{D} \ \text{ s.t. } \ \mathrm{E}(\zeta) < x\right\}$$

and $$\mathcal{L}_{\inf}^{>}(x,\nu) = \inf\left\{\mathsf{KL}(\zeta,\nu) : \zeta \in \mathcal{D} \ \text{ s.t. } \ \mathrm{E}(\zeta) > x\right\}$$

Non-parametric upper bound

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \leq -\frac{1}{\overline{\ln}\, K} \min_{2\leq k\leq K} \frac{\mathcal{L}\big(\nu_{\sigma_k}, \nu^\star\big)}{k}$$

where arms are ordered such that

$$0 = \mathcal{L}\big(\nu_{\sigma_1}, \nu^\star\big) < \mathcal{L}\big(\nu_{\sigma_2}, \nu^\star\big) \leq \ldots \leq \mathcal{L}\big(\nu_{\sigma_{K-1}}, \nu^\star\big) \leq \mathcal{L}\big(\nu_{\sigma_K}, \nu^\star\big)$$

Pinsker's inequality: $\mathcal{L}(\nu_2, \nu_1) \geq (\mu_1 - \mu_2)^2$

$\rightarrow$ Generalization and improvement of the bound of [Audibert et al., 2010]

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \leq -\frac{1}{\overline{\ln}\, K} \min_{2\leq k\leq K} \frac{\Delta_{(k)}^2}{k}$$

Non-parametric lower bounds

Recipe For any bandit problem $\underline{\lambda}$ such that $a^\star(\underline{\lambda}) \neq a^\star(\underline{\nu})$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \geq -\limsup_{T\to+\infty} \sum_{a=1}^{K} \frac{\mathbb{E}_{\underline{\lambda}}[N_a(T)]}{T} \mathrm{KL}(\lambda_a, \nu_a)$$

Difficulty Control of $\frac{\mathbb{E}_{\underline{\lambda}}[N_a(T)]}{T}$ when $\lambda_a \neq \nu_a$.

Gap-based bound for a Bernoulli model [Audibert et al., 2010]

$\rightarrow$ **Careful construction of $\underline{\lambda}$ with consistency assumption**

Adaptation to non-parametric models?

$\rightarrow$ **Natural hypothesis on the strategies**

Weak monotonicity

$$\limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\nu}}[N_{(K)}(T)]}{T} \leq \frac{1}{K}$$

Strong monotonicity

$$\limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\nu}}[N_{(a)}(T)]}{T} \leq \frac{1}{a}$$

for all $a \in [K]$

$\rightarrow$ satisfied by the successive-rejects strategy

Non-parametric lower bounds

For a general model $\mathcal{D}$, under some assumptions including **weak** monotonicity

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \geq -\min_{2\leq k\leq K} \frac{\mathcal{L}^{<}_{\inf}\big(\mu_{(k)}, \nu^\star\big)}{k}$$

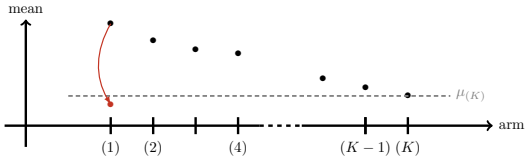
Sketch of proof (case $k = K$)

① Weak monotonicity on $\underline{\lambda}$ only differing from $\underline{\nu}$ at arm (1) with $\mathrm{E}(\lambda_{(1)}) < \mu_{(K)}$

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \geq -\limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\lambda}}[N_{(1)}(T)]}{T} \mathrm{KL}(\lambda_{(1)}, \nu_{(1)}) \geq -\frac{\mathrm{KL}(\lambda_{(1)}, \nu^\star)}{K}$$

② Take infimum on $\mathrm{E}(\lambda_{(1)}) < \mu_{(K)}$

$$\inf_{\lambda_{(1)}\in\mathcal{D} \text{ s.t. } \mathrm{E}(\lambda_{(1)})<\mu_{(K)}} \mathrm{KL}(\lambda_{(1)}, \nu^\star) = \mathcal{L}^{<}_{\inf}(\mu_{(K)}, \nu^\star)$$



Non-parametric lower bounds

$\rightarrow$ Generalization and improvement of the bound of [Audibert et al., 2010]

For the model $\mathcal{D} = \mathcal{B}_{[p,\,1-p]} = \big\{\mathrm{Ber}(x) : x \in [p, 1-p]\big\}$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \geq -\frac{5}{p(1-p)} \min_{2\leq k\leq K} \frac{\Delta_{(k)}^2}{k}$$

Improvement towards cutoff $x^\star$

For a general model $\mathcal{D}$ and a consistent and **strongly** monotonous strategy

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(\hat{a}_T \neq a^\star(\underline{\nu})\big) \geq -\min_{2\leq k\leq K} \inf_{x\in[\mu_{(k)},\mu_{(k-1)})} \frac{\mathcal{L}^{>}_{\inf}\big(x, \nu_{(k)}\big) + \mathcal{L}^{<}_{\inf}\big(x, \nu^\star\big)}{k-1}$$