Solving pure exploration problems with the Top Two approach

Marc Jourdan

Supervised by Dr. **Émilie Kaufmann** and Dr. **Rémy Degenne**

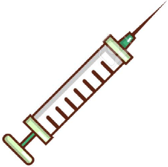
June 14, 2024

Phase III clinical trials

μ_1

μ_2

μ_3



μ_4

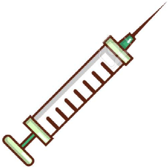
Goal: Identify a treatment with a high efficiency.

Phase III clinical trials

μ_1

μ_2

μ_3



μ_4

Goal: Identify a treatment with a high efficiency.

Setting: Pure exploration for stochastic multi-armed bandits.

☞ Sequential hypothesis testing with adaptive data collection.

Sequential decision making under uncertainty

After treating $n-1$ patients, the physician has

☞ a guessed answer for a good treatment $\hat{\imath}_n \in [K]$.

As the n-th patient enters, the physician selects

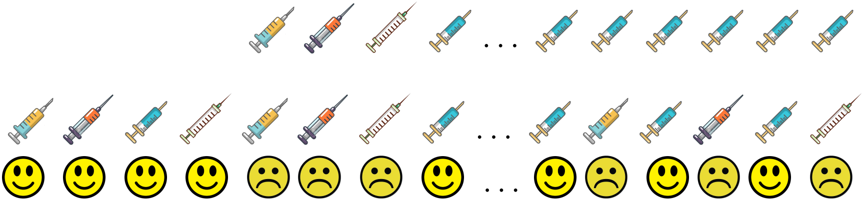
☞ a treatment $I_n \in [K]$ for administration.

Then, it observes a realization $X_n \sim \nu_{I_n}$ with $\nu_i = \mathcal{B}(\mu_i)$.

$(\hat{\imath}_n)_{n>K}$

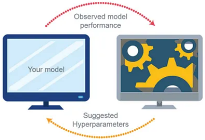
$(I_n)_{n\geq 1}$

$(X_n)_{n\geq 1}$



Other applications

- crop management for agriculture,
- A/B testing for online marketing,
- hyperparameter optimization.



Key requirements of a good strategy

To be advocated by statisticians:

✓ guarantees on the quality of the guessed answer,

✓ low empirical error quickly.

Key requirements of a good strategy

To be advocated by statisticians:

✓ guarantees on the quality of the guessed answer,

✓ low empirical error quickly.

To be used by practitioners:

✓ simple,

✓ interpretable,

✓ generalizable,

✓ versatile.

Key requirements of a good strategy

To be advocated by statisticians:

✓ guarantees on the quality of the guessed answer,

✓ low empirical error quickly.

To be used by practitioners:

✓ simple,

✓ interpretable,

✓ generalizable,

✓ versatile.

The **Top Two approach** satisfies them all !

The Top Two approach

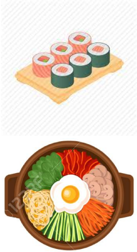
Set a **leader** answer $B_n \in [K]$;

Set a **challenger** answer $C_n \in [K] \setminus \{B_n\}$;

Set a **target** $\beta_n(B_n, C_n) \in [0, 1]$;

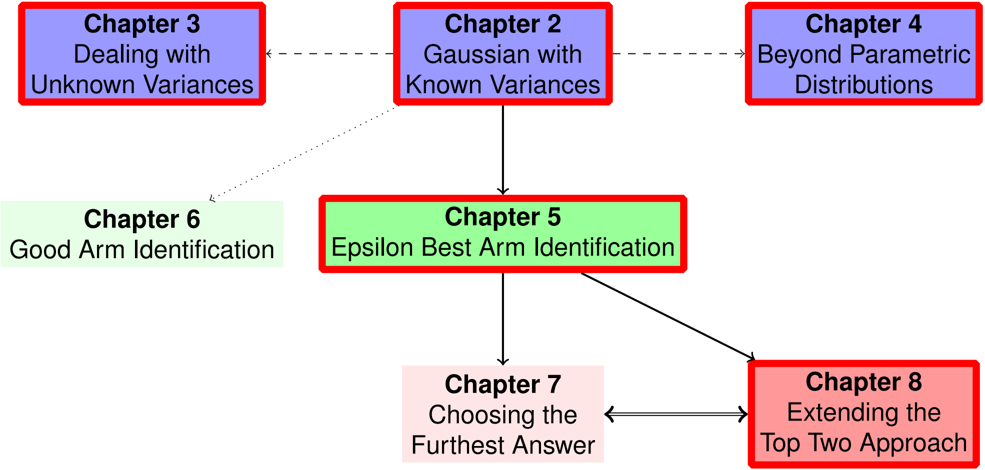
Return $I_n \in \{B_n, C_n\}$ using target $\beta_n(B_n, C_n)$.

Answers



Target

Roadmap of this talk based on my PhD thesis



Contributions featured in this talk

- **MJ**, Rémy Degenne, Dorian Baudry, Rianne de Heide and Émilie Kaufmann. Top Two algorithms revisited.
  Advances in Neural Information Processing Systems, 2022.
- **MJ**, Rémy Degenne and Émilie Kaufmann. Dealing with unknown variances in best-arm identification.
  Algorithmic Learning Theory, 2023.
- **MJ** and Rémy Degenne. Non-asymptotic analysis of a UCB-based Top Two algorithm.
  Advances in Neural Information Processing Systems, 2023.
- **MJ**, Rémy Degenne and Émilie Kaufmann. An ε-best-arm identification algorithm for fixed-confidence and beyond.
  Advances in Neural Information Processing Systems, 2023.

Other contributions during my PhD thesis

- **MJ** and Rémy Degenne. Choosing answers in ε-best-answer identification for linear bandits.
  International Conference on Machine Learning, 2022.
- Achraf Azize, **MJ**, Aymen Al Marjani and Debabrota Basu. On the complexity of differentially private best-arm identification with fixed confidence.
  Advances in Neural Information Processing Systems, 2023.
- **MJ** and Clémence Réda. An anytime algorithm for good arm identification.
- Achraf Azize, **MJ**, Aymen Al Marjani and Debabrota Basu. Differentially private best-arm identification.

Stochastic multi-armed bandits

K arms: arm $i \in [K]$ with $\nu_i \in \mathcal{D}$ having mean μ_i .

Class of distributions $\mathcal{D}$:

- parametric, e.g. Bernoulli, Gaussian (known or unknown variance).
- non-parametric, e.g. bounded distributions in $[0, B]$.

Stochastic multi-armed bandits

K arms: arm $i \in [K]$ with $\nu_i \in \mathcal{D}$ having mean μ_i .

Class of distributions $\mathcal{D}$:

- parametric, e.g. Bernoulli, Gaussian (known or unknown variance).
- non-parametric, e.g. bounded distributions in $[0, B]$.

Underlying structure:

- vanilla, $\mu = (\mu_i)_{i \in [K]} \in \mathbb{R}^K$.
- linear, $\mu_i = \langle \theta, a_i \rangle$ where $\theta \in \mathbb{R}^d$ is unknown and $a_i \in \mathbb{R}^d$ is known.

Stochastic multi-armed bandits

K arms: arm $i \in [K]$ with $\nu_i \in \mathcal{D}$ having mean μ_i .

Class of distributions $\mathcal{D}$:

- parametric, e.g. Bernoulli, Gaussian (known or unknown variance).
- non-parametric, e.g. bounded distributions in $[0, B]$.

Underlying structure:

- vanilla, $\mu = (\mu_i)_{i \in [K]} \in \mathbb{R}^K$.
- linear, $\mu_i = \langle \theta, a_i \rangle$ where $\theta \in \mathbb{R}^d$ is unknown and $a_i \in \mathbb{R}^d$ is known.

Running example

Vanilla bandits for **Gaussian** with unit variance.

ε-Best-Arm Identification (ε-BAI)

Goal: identify one arm in $\mathcal{I}_\varepsilon(\mu) = \{i \mid \mu_i \geq \max_j \mu_j - \varepsilon\}$ with $\varepsilon \geq 0$.

Algorithm: at time n ,

- *Recommendation rule*: recommend a candidate answer $\hat{\imath}_n$.
- *Stopping rule* (optional): dictate when to stop sampling .
- **Sampling rule**: pull an arm I_n and observe $X_n \sim \nu_{I_n}$.

Performance metrics

Fixed-confidence: given an error/confidence pair (ε, δ) ,

- Define an (ε, δ)-PAC stopping time $\tau_{\varepsilon,\delta}$, i.e.

$$\mathbb{P}_\nu(\tau_{\varepsilon,\delta} < +\infty, \hat{\imath}_{\tau_{\varepsilon,\delta}} \notin \mathcal{I}_\varepsilon(\mu)) \leq \delta \ .$$

- Minimize the **expected sample complexity** $\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]$.

Performance metrics

Fixed-confidence: given an error/confidence pair (ε, δ) ,

☞ Define an (ε, δ)-PAC stopping time $\tau_{\varepsilon,\delta}$, i.e.

$$\mathbb{P}_\nu(\tau_{\varepsilon,\delta} < +\infty, \hat{\imath}_{\tau_{\varepsilon,\delta}} \notin \mathcal{I}_\varepsilon(\mu)) \leq \delta \, .$$

☞ Minimize the **expected sample complexity** $\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]$.

Fixed-budget: given an error/budget pair (ε, T) ,

☞ Minimize the **probablity of ε-error** $\mathbb{P}_\nu(\hat{\imath}_T \notin \mathcal{I}_\varepsilon(\mu))$ at time T .

Performance metrics

Fixed-confidence: given an error/confidence pair (ε, δ) ,

- ☞ Define an (ε, δ)-PAC stopping time $\tau_{\varepsilon,\delta}$, i.e.

$$\mathbb{P}_\nu(\tau_{\varepsilon,\delta} < +\infty, \hat{\imath}_{\tau_{\varepsilon,\delta}} \notin \mathcal{I}_\varepsilon(\mu)) \leq \delta \, .$$

- ☞ Minimize the **expected sample complexity** $\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]$.

Fixed-budget: given an error/budget pair (ε, T) ,

- ☞ Minimize the **probablity of ε-error** $\mathbb{P}_\nu(\hat{\imath}_T \notin \mathcal{I}_\varepsilon(\mu))$ at time T .

Anytime: Control the **simple regret** $\mathbb{E}_\nu[\max_j \mu_j - \mu_{\hat{\imath}_n}]$ at any time n.

Lower bound on the expected sample complexity

(Garivier and Kaufmann, 2016; Degenne and Koolen, 2019; Agrawal et al., 2020)

For all (ε, δ)-PAC algorithm and all instances $\nu \in \mathcal{D}^K$,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]}{\log(1/\delta)} \geq T_\varepsilon(\nu) \; ,$$

Lower bound on the expected sample complexity

(Garivier and Kaufmann, 2016; Degenne and Koolen, 2019; Agrawal et al., 2020)

For all (ε, δ)-PAC algorithm and all instances $\nu \in \mathcal{D}^K$,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]}{\log(1/\delta)} \geq T_\varepsilon(\nu) ,$$

where the inverse of the characteristic time is

$$T_\varepsilon(\nu)^{-1} = \max_{i \in \mathcal{I}_\varepsilon(\mu)} \max_{w \in \triangle_K} \min_{j \neq i} C_\varepsilon(i, j; \nu, w) ,$$

reached at the optimal allocation $w_\varepsilon(\nu)$ and furthest answer $i_F(\nu)$.

Lower bound on the expected sample complexity

(Garivier and Kaufmann, 2016; Degenne and Koolen, 2019; Agrawal et al., 2020)

For all (ε, δ)-PAC algorithm and all instances $\nu \in \mathcal{D}^K$,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]}{\log(1/\delta)} \geq T_\varepsilon(\nu) ,$$

where the inverse of the characteristic time is

$$T_\varepsilon(\nu)^{-1} = \max_{i \in \mathcal{I}_\varepsilon(\mu)} \max_{w \in \triangle_K} \min_{j \neq i} C_\varepsilon(i, j; \nu, w) ,$$

reached at the optimal allocation $w_\varepsilon(\nu)$ and furthest answer $i_F(\nu)$.

Vanilla bandits for Gaussian with unit variance

$$C_\varepsilon(i, j; \nu, w) = \mathbb{1}\left(\mu_i > \mu_j - \varepsilon\right) \frac{(\mu_i - \mu_j + \varepsilon)^2}{2(1/w_i + 1/w_j)} .$$

How to obtain an (ε, δ)-PAC algorithm ?

☞ recommend the empirical best arm

$$\hat{i}_n = \arg\max_{i \in [K]} \mu_{n,i} \ ,$$

with $\mu_{n,i} = N_{n,i}^{-1} \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right) X_t$ and $N_{n,i} = \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right)$.

How to obtain an (ε, δ)-PAC algorithm ?

- recommend the empirical best arm

$$\hat{i}_n = \arg\max_{i \in [K]} \mu_{n,i} \ ,$$

with $\mu_{n,i} = N_{n,i}^{-1} \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right) X_t$ and $N_{n,i} = \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right)$.

- Generalized likelihood ratio (**GLR**) stopping rule

$$\tau_{\varepsilon,\delta} = \inf\{n \in \mathbb{N} \mid \min_{j \neq \hat{i}_n} C_{\varepsilon,n}(\hat{i}_n, j) > c(n-1, \delta)\} \ ,$$

with $C_{\varepsilon,n}(i,j) = C_\varepsilon(i,j;\nu_n, N_n)$ and $c(n,\delta) \approx \log(1/\delta) + \mathcal{O}(\log n)$.

How to obtain an (ε, δ)-PAC algorithm ?

- recommend the empirical best arm

$$\hat{i}_n = \arg\max_{i \in [K]} \mu_{n,i} \ ,$$

with $\mu_{n,i} = N_{n,i}^{-1} \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right) X_t$ and $N_{n,i} = \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right)$.

- Generalized likelihood ratio (**GLR**) stopping rule

$$\tau_{\varepsilon,\delta} = \inf\{n \in \mathbb{N} \mid \min_{j \neq \hat{i}_n} C_{\varepsilon,n}(\hat{i}_n, j) > c(n-1, \delta)\} \ ,$$

with $C_{\varepsilon,n}(i,j) = C_{\varepsilon}(i,j;\nu_n, N_n)$ and $c(n,\delta) \approx \log(1/\delta) + \mathcal{O}(\log n)$.

Vanilla bandits for Gaussian with unit variance

$$C_{\varepsilon,n}(i,j) = \mathbb{1}\left(\mu_{n,i} > \mu_{n,j} - \varepsilon\right) \frac{(\mu_{n,i} - \mu_{n,j} + \varepsilon)^2}{2(1/N_{n,i} + 1/N_{n,j})} \ .$$

Lower bound based sampling rules

Track-and-Stop (Garivier and Kaufmann, 2016) At n, solve $w_n = w_\varepsilon(\nu_n)$.

Lower bound based sampling rules

Track-and-Stop (Garivier and Kaufmann, 2016)

At n, solve $w_n = w_\varepsilon(\nu_n)$.

Online optimization approach:

- DKM (Degenne et al., 2019),
- FWS (Wang et al., 2021).

At n, get w_n from learner $\mathcal{L}^K$;
Feed loss $\ell_n(w)$ to learner $\mathcal{L}^K$.

Lower bound based sampling rules

Track-and-Stop (Garivier and Kaufmann, 2016)

At n, solve $w_n = w_\varepsilon(\nu_n)$.

Online optimization approach:

- DKM (Degenne et al., 2019),
- FWS (Wang et al., 2021).

At n, get w_n from learner $\mathcal{L}^K$;
Feed loss $\ell_n(w)$ to learner $\mathcal{L}^K$.

Top Two approach:

- LUCB (Kalyanakrishnan et al., 2012),
- TTTS (Russo, 2016),
- TTEI (Qin et al., 2017),
- T3C (Shang et al., 2020).

At n, set leader answer B_n ;
Set challenger answer $C_n \neq B_n$;
Set target $\beta_n(B_n, C_n) \in [0, 1]$;
Set $I_n \in \{B_n, C_n\}$ with $\beta_n(B_n, C_n)$.

The greedy GLR-based sampling rule

At time $n < \tau_{\varepsilon,\delta}$,

- **candidate** (or **leader**) answer, $\hat{\imath}_n = \arg\max_{i \in [K]} \mu_{n,i}$,
- **alternative** (or **challenger**) answer, $\hat{\jmath}_n = \arg\min_{j \neq \hat{\imath}_n} C_{\varepsilon,n}(\hat{\imath}_n, j)$.

The greedy GLR-based sampling rule

At time $n < \tau_{\varepsilon,\delta}$,

- **candidate** (or **leader**) answer, $\hat{\imath}_n = \arg\max_{i \in [K]} \mu_{n,i}$,
- **alternative** (or **challenger**) answer, $\hat{\jmath}_n = \arg\min_{j \neq \hat{\imath}_n} C_{\varepsilon,n}(\hat{\imath}_n, j)$.

Since we don't stop, i.e. $C_{\varepsilon,n}(\hat{\imath}_n, \hat{\jmath}_n) \leq c(n-1, \delta)$, we want to

☞ verify that $\hat{\imath}_n$ is better than $\hat{\jmath}_n$,

☞ hence we sample $I_n \in \{\hat{\imath}_n, \hat{\jmath}_n\}$.

The greedy GLR-based sampling rule

At time $n < \tau_{\varepsilon,\delta}$,

- **candidate** (or **leader**) answer, $\hat{\imath}_n = \arg\max_{i \in [K]} \mu_{n,i}$,
- **alternative** (or **challenger**) answer, $\hat{\jmath}_n = \arg\min_{j \neq \hat{\imath}_n} C_{\varepsilon,n}(\hat{\imath}_n, j)$.

Since we don't stop, i.e. $C_{\varepsilon,n}(\hat{\imath}_n, \hat{\jmath}_n) \leq c(n-1, \delta)$, we want to

☞ verify that $\hat{\imath}_n$ is better than $\hat{\jmath}_n$,

☞ hence we sample $I_n \in \{\hat{\imath}_n, \hat{\jmath}_n\}$.

⚠ When ε is small, this is too greedy in practice.

☞ Implicit exploration when selecting $\hat{\imath}_n$ or $\hat{\jmath}_n$.

The Top Two approach

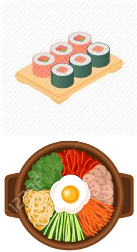
Set a **leader** answer $B_n \in [K]$;

Set a **challenger** answer $C_n \in [K] \setminus \{B_n\}$;

Set a **target** $\beta_n(B_n, C_n) \in [0, 1]$;

Return $I_n \in \{B_n, C_n\}$ using target $\beta_n(B_n, C_n)$.

Answers



Target

Leader answer $B_n \in [K]$

☞ **Empirical Best** (EB) (Jourdan et al., 2022), $\arg\max_{i \in [K]} \mu_{n,i}$.

Leader answer $B_n \in [K]$

- **Empirical Best** (EB) (Jourdan et al., 2022), $\arg\max_{i \in [K]} \mu_{n,i}$.

- **Upper Confidence Bound** (UCB) (Jourdan and Degenne, 2023),

$$\arg\max_{i \in [K]} U_{n,i} \quad \text{with} \quad U_{n,i} = \arg\max \left\{ \lambda \mid N_{n,i} \mathrm{KL}(\mu_{n,i}, \lambda) \lesssim \log(n) \right\} .$$

Leader answer $B_n \in [K]$

- **Empirical Best** (EB) (Jourdan et al., 2022), $\arg\max_{i \in [K]} \mu_{n,i}$.
- **Upper Confidence Bound** (UCB) (Jourdan and Degenne, 2023),

$$\arg\max_{i \in [K]} U_{n,i} \quad \text{with} \quad U_{n,i} = \arg\max \{\lambda \mid N_{n,i} \mathrm{KL}(\mu_{n,i}, \lambda) \lesssim \log(n)\} \ .$$

- Thompson Sampling (TS) (Russo, 2016),

$$\arg\max_{i \in [K]} \theta_{n,i} \quad \text{with} \quad \theta_n \sim \Pi_n = \bigotimes_{i \in [K]} \Pi_{n,i} \ .$$

Leader answer $B_n \in [K]$

- **Empirical Best** (EB) (Jourdan et al., 2022), $\arg\max_{i \in [K]} \mu_{n,i}$.
- **Upper Confidence Bound** (UCB) (Jourdan and Degenne, 2023),

$$\arg\max_{i \in [K]} U_{n,i} \quad \text{with} \quad U_{n,i} = \arg\max \{\lambda \mid N_{n,i} \mathrm{KL}(\mu_{n,i}, \lambda) \lesssim \log(n)\} \ .$$

- Thompson Sampling (TS) (Russo, 2016),

$$\arg\max_{i \in [K]} \theta_{n,i} \quad \text{with} \quad \theta_n \sim \Pi_n = \bigotimes_{i \in [K]} \Pi_{n,i} \ .$$

Vanilla bandits for Gaussian with unit variance

$$U_{n,i} \approx \mu_{n,i} + \sqrt{2 \log(n) / N_{n,i}} \quad \text{and} \quad \Pi_{n,i} = \mathcal{N}(\mu_{n,i}, 1/N_{n,i}) \ .$$

Challenger answer $C_n \in [K] \setminus \{B_n\}$

☞ Transportation Cost (TC) (Shang et al., 2020),

$$\arg\min_{j \neq B_n} C_{\varepsilon,n}(B_n, j) \ .$$

Challenger answer $C_n \in [K] \setminus \{B_n\}$

- Transportation Cost (TC) (Shang et al., 2020),

$$\arg\min_{j \neq B_n} C_{\varepsilon,n}(B_n, j) \; .$$

- **Transportation Cost Improved** (TCI) (Jourdan et al., 2022),

$$\arg\min_{j \neq B_n} \left\{ C_{\varepsilon,n}(B_n, j) + \log N_{n,j} \right\} \; .$$

Challenger answer $C_n \in [K] \setminus \{B_n\}$

- Transportation Cost (TC) (Shang et al., 2020),

$$\arg\min_{j \neq B_n} C_{\varepsilon,n}(B_n, j) \, .$$

- **Transportation Cost Improved** (TCI) (Jourdan et al., 2022),

$$\arg\min_{j \neq B_n} \{C_{\varepsilon,n}(B_n, j) + \log N_{n,j}\} \, .$$

- Re-Sampling (RS) (Russo, 2016),

$$\arg\max_{i \in [K]} \theta_{n,i} \quad \text{with} \quad \theta_n \sim \Pi_n \quad \text{until} \quad B_n \notin \mathcal{I}_\varepsilon(\theta_n) \, .$$

Target allocation $\beta_n(B_n, C_n) \in [0, 1]$

☞ **Fixed** design (Russo, 2016),

$$\beta_n(i, j) = \beta \in (0, 1) \; .$$

Target allocation $\beta_n(B_n, C_n) \in [0, 1]$

☞ **Fixed** design (Russo, 2016),

$$\beta_n(i, j) = \beta \in (0, 1) \ .$$

☞ Optimal design **IDS** (Information Directed Selection) (You et al., 2023),

$$\beta_n(i, j) = \frac{N_{n,i}}{C_{\varepsilon,n}(i, j)} \frac{\partial C_\varepsilon}{\partial w_i}(i, j; \nu_n, N_n) \ ,$$

when $\mu_{n,i} > \mu_{n,j} - \varepsilon$, and $\beta_n(i, j) = 1/2$ otherwise.

Target allocation $\beta_n(B_n, C_n) \in [0, 1]$

☞ **Fixed** design (Russo, 2016),

$$\beta_n(i, j) = \beta \in (0, 1) .$$

☞ Optimal design **IDS** (Information Directed Selection) (You et al., 2023),

$$\beta_n(i, j) = \frac{N_{n,i}}{C_{\varepsilon,n}(i, j)} \frac{\partial C_\varepsilon}{\partial w_i}(i, j; \nu_n, N_n) ,$$

when $\mu_{n,i} > \mu_{n,j} - \varepsilon$, and $\beta_n(i, j) = 1/2$ otherwise.

Vanilla bandits for Gaussian with unit variance

When $\mu_{n,i} > \mu_{n,j} - \varepsilon, \quad \beta_n(i, j) = N_{n,j}/(N_{n,i} + N_{n,j}) .$

Reaching the target

☞ Randomized (Russo, 2016),

$$I_n = \begin{cases} B_n & \text{with probability } \beta_n(B_n, C_n) \ , \\ C_n & \text{otherwise} \ . \end{cases}$$

Reaching the target

☞ Randomized (Russo, 2016),

$$I_n = \begin{cases} B_n & \text{with probability } \beta_n(B_n, C_n) \, , \\ C_n & \text{otherwise} \, . \end{cases}$$

☞ **Tracking** (Jourdan and Degenne, 2023),

$$I_n = \begin{cases} C_n & \text{if } N^{B_n}_{n,C_n} \leq (1 - \overline{\beta}_{n+1}(B_n, C_n)) T_{n+1}(B_n, C_n) \, , \\ B_n & \text{otherwise} \, . \end{cases}$$

with $N^i_{n,j} = \sum_{t \in [n-1]} \mathbb{1}\left((B_t, C_t) = (i,j), I_t = j\right)$, $T_n(i,j) = \sum_{t \in [n-1]} \mathbb{1}\left((B_t, C_t) = (i,j)\right)$ and $\overline{\beta}_n(i,j) = T_n(i,j)^{-1} \sum_{t \in [n-1]} \beta_t(i,j) \mathbb{1}\left((B_t, C_t) = (i,j)\right)$.

Asymptotic (β-)optimality

Theorem (Jourdan et al. 2022; Jourdan and Degenne 2023; Jourdan et al. 2023a)

The Top Two sampling rule with any pair of leader/challenger satisfying some properties yields an (ε, δ)-PAC algorithm and, for all $\nu \in \mathcal{D}^K$ with unique best arm (and distinct means for $\varepsilon = 0$),

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]}{\log(1/\delta)} \leq \begin{cases} T_\varepsilon(\nu) & \textbf{[IDS]} \\ T_{\varepsilon,\beta}(\nu) & \textbf{[fixed } \beta \textbf{]} \end{cases} \quad \textit{with} \quad T_{\varepsilon,1/2}(\nu) \leq 2T_\varepsilon(\nu) \, .$$

Asymptotic (β-)optimality

Theorem (Jourdan et al. 2022; Jourdan and Degenne 2023; Jourdan et al. 2023a)

The Top Two sampling rule with any pair of leader/challenger satisfying some properties yields an (ε, δ)-PAC algorithm and, for all $\nu \in \mathcal{D}^K$ with unique best arm (and distinct means for $\varepsilon = 0$),

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_{\varepsilon,\delta}]}{\log(1/\delta)} \le \begin{cases} T_\varepsilon(\nu) & \textbf{[IDS]} \\ T_{\varepsilon,\beta}(\nu) & \textbf{[fixed } \beta\textbf{]} \end{cases} \quad \textit{with} \quad T_{\varepsilon,1/2}(\nu) \le 2T_\varepsilon(\nu) \, .$$

Distributions $\mathcal{D}$	IDS	Fixed	TS	EB	UCB	RS	TC	TCI
Gaussian KV	✓	✓	✓	✓	✓	✓	✓	✓
(Shang et al., 2020; You et al., 2023)	YQWY23	SdHK+20	SdHK+20	JD+22	JD23	SdHK+20	SdHK+20	JD+22
Bernoulli	?	✓	✓	✓	✓	✓	✓	✓
sub-Exp 1-Exp.Fam.	?	✓	?	✓	✓	?	✓	✓
Gaussian UV	?	✓	?	✓	✓	?	✓	✓
Bounded	?	✓	✓	✓	✓	✓	✓	✓

Beyond Gaussian with unit variance

Empirical transportation cost for a class of distributions $\mathcal{D}$,

$$C_{\varepsilon,n}(i,j) = \mathbb{1}\left(\mu_{n,i} > \mu_{n,j} - \varepsilon\right) \inf_{u \in \mathcal{I}} \left\{ N_{n,i} \mathcal{K}_{\inf}^{-}(\nu_{n,i}, u - \varepsilon) + N_{n,j} \mathcal{K}_{\inf}^{+}(\nu_{n,j}, u) \right\} ,$$

where $\mathcal{K}_{\inf}^{+}(\nu, u) = \inf\{\mathrm{KL}(\nu, \kappa) \mid \kappa \in \mathcal{D},\ \mathbb{E}_{X \sim \kappa}[X] > u\}$.

Beyond Gaussian with unit variance

Empirical transportation cost for a class of distributions $\mathcal{D}$,

$$C_{\varepsilon,n}(i,j) = \mathbb{1}\left(\mu_{n,i} > \mu_{n,j} - \varepsilon\right) \inf_{u \in \mathcal{I}} \left\{ N_{n,i} \mathcal{K}^{-}_{\inf}(\nu_{n,i}, u - \varepsilon) + N_{n,j} \mathcal{K}^{+}_{\inf}(\nu_{n,j}, u) \right\} ,$$

where $\mathcal{K}^{+}_{\inf}(\nu, u) = \inf\{\mathrm{KL}(\nu, \kappa) \mid \kappa \in \mathcal{D},\ \mathbb{E}_{X\sim\kappa}[X] > u\}$.

- ☞ Gaussian with **unknown variance**,

$$\mathcal{K}^{+}_{\inf}(\nu_{n,i}, u) = \mathbb{1}\left(\mu_{n,i} < u\right) \frac{1}{2} \log\left(1 + \frac{(\mu_{n,i} - u)^2}{\sigma^2_{n,i}}\right) ,$$

where $\mathcal{I} = \mathbb{R}$ and $\sigma^2_{n,i} = N_{n,i}^{-1} \sum_{t \in [n-1]} \mathbb{1}\left(I_t = i\right) (X_t - \mu_{n,i})^2$.

- ☞ **Bounded** distributions with known support $\mathcal{I} = [0, B]$.

Proof sketch

❶ Let T_γ such that $\max_{i \neq i^\star} \left| \frac{N_{n,i}}{N_{n,i^\star}} - \frac{w_{\varepsilon,i}}{w_{\varepsilon,i^\star}} \right| \leq \gamma$ for all $n \geq T_\gamma$.

$$\log(1/\delta) \approx_{\delta \to 0} c(n, \delta) \geq \min_{j \neq \hat{\imath}_n} C_{\varepsilon,n}(\hat{\imath}_n, j) \approx_{n \geq T_\gamma} n T_\varepsilon(\nu)^{-1} .$$

Proof sketch

❶ Let T_γ such that $\max_{i \neq i^\star} \left| \frac{N_{n,i}}{N_{n,i^\star}} - \frac{w_{\varepsilon,i}}{w_{\varepsilon,i^\star}} \right| \leq \gamma$ for all $n \geq T_\gamma$.

$$\log(1/\delta) \approx_{\delta \to 0} c(n, \delta) \geq \min_{j \neq \hat{\imath}_n} C_{\varepsilon,n}(\hat{\imath}_n, j) \approx_{n \geq T_\gamma} n T_\varepsilon(\nu)^{-1} .$$

❷ **Sufficient exploration**, i.e. $\min_{i \in [K]} N_{n,i} \geq \sqrt{n/K}$ for n large.

If there are undersampled arms, then either the leader or the challenger is one of them. As it will be sampled, this yields a contradiction.

Proof sketch

❶ Let T_γ such that $\max_{i \neq i^\star} \left| \frac{N_{n,i}}{N_{n,i^\star}} - \frac{w_{\varepsilon,i}}{w_{\varepsilon,i^\star}} \right| \leq \gamma$ for all $n \geq T_\gamma$.

$$\log(1/\delta) \approx_{\delta \to 0} c(n,\delta) \geq \min_{j \neq \hat{\imath}_n} C_{\varepsilon,n}(\hat{\imath}_n, j) \approx_{n \geq T_\gamma} n T_\varepsilon(\nu)^{-1} .$$

❷ **Sufficient exploration**, i.e. $\min_{i \in [K]} N_{n,i} \geq \sqrt{n/K}$ for n large.

If there are undersampled arms, then either the leader or the challenger is one of them. As it will be sampled, this yields a contradiction.

❸ **Convergence towards $\boldsymbol{w_\varepsilon(\nu)}$**, i.e. $\mathbb{E}_\nu[T_\gamma] < +\infty$ for γ small.

If an arm overshoots the ratio of optimal allocation with $i^\star$, then it will not be chosen as challenger. Therefore, the ratio will converge.

Input: **slack** $\varepsilon > 0$, proportion $\beta \in (0, 1)$ (only for fixed).

Set $\hat{\imath}_n \in \arg\max_{i \in [K]} \mu_{n,i}$;
Set $B_n = \hat{\imath}_n$;
Set $C_n \in \arg\min_{i \neq B_n} \frac{\mu_{n,B_n} - \mu_{n,i} + \varepsilon}{\sqrt{1/N_{n,B_n} + 1/N_{n,i}}}$;

Set $\overline{\beta}_{n+1}(B_n, C_n)$ with $\beta_n(i,j) = \begin{cases} \beta & \text{[fixed]} \\ \frac{N_{n,j}}{N_{n,i} + N_{n,j}} & \text{[IDS]} \end{cases}$;

Set $I_n = \begin{cases} C_n & \text{if } N^{B_n}_{n,C_n} \leq (1 - \overline{\beta}_{n+1}(B_n, C_n)) T_{n+1}(B_n, C_n) \, , \\ B_n & \text{otherwise} \, . \end{cases}$

Output: next arm to sample I_n and next recommendation $\hat{\imath}_n$.

Expected sample complexity

Theorem (Jourdan et al. 2023b)

EB-TC$_\varepsilon$ with IDS (resp. fixed β) proportions is (ε, δ)-PAC and **asymptotically** *(resp. β-)***optimal** *for ε-BAI on instances with unique best arm.*

Expected sample complexity

Theorem (Jourdan et al. 2023b)

EB-TC$_\varepsilon$ with IDS (resp. fixed β) proportions is (ε, δ)-PAC and **asymptotically** *(resp. β-)***optimal** *for ε-BAI on instances with unique best arm.*
On any instances, EB-TC$_\varepsilon$ with fixed $\beta = 1/2$ satisfies that

$$\mathbb{E}_\nu[\tau_{\varepsilon,\delta}] \leq \inf_{x \in [0,\varepsilon]} \max \{T_{\nu,\varepsilon}(\delta, x) + 1, \ S_{\nu,\varepsilon}(x)\} + 2K^2 \,, \quad \textit{where}$$

$\lim_{\delta \to 0} \frac{T_{\mu,\varepsilon}(\delta,0)}{\log(1/\delta)} \leq 2|i^\star(\mu)| T_{\varepsilon,1/2}(\nu)$, $S_{\nu,\varepsilon}(\varepsilon/2) = \mathcal{O}(K^2 |\mathcal{I}_{\varepsilon/2}(\mu)| \varepsilon^{-2} \log \varepsilon^{-1})$.

Any time and uniform probability of ε-error

Theorem (Jourdan et al. 2023b)

EB-TC$_\varepsilon$ with fixed $\beta = 1/2$ satisfies that, for all $n > 5K^2/2$ and all $\tilde{\varepsilon} \geq 0$,

$$\mathbb{P}_\nu\left(\hat{\imath}_n \notin \mathcal{I}_{\tilde{\varepsilon}}(\mu)\right) \leq \exp\left(-\Theta\left(\frac{n}{H_{i_\mu(\tilde{\varepsilon})}(\mu, \varepsilon)}\right)\right) ,$$

where $H_1(\mu, \varepsilon) = K(2\Delta_{\min}^{-1} + 3\varepsilon^{-1})^2$ and $H_i(\mu, \varepsilon) = \Theta(K/\Delta_{i+1}^{-2})$. Ordered distinct mean gaps $(\Delta_i)_{i \in [C_\mu]}$ and $i_\mu(\tilde{\varepsilon}) = i$ if $\tilde{\varepsilon} \in [\Delta_i, \Delta_{i+1})$.

Any time and uniform probability of ε-error

Theorem (Jourdan et al. 2023b)

EB-TC$_\varepsilon$ with fixed $\beta = 1/2$ satisfies that, for all $n > 5K^2/2$ and all $\tilde{\varepsilon} \geq 0$,

$$\mathbb{P}_\nu \left(\hat{\imath}_n \notin \mathcal{I}_{\tilde{\varepsilon}}(\mu) \right) \leq \exp \left(-\Theta \left(\frac{n}{H_{i_\mu(\tilde{\varepsilon})}(\mu, \varepsilon)} \right) \right) ,$$

where $H_1(\mu, \varepsilon) = K(2\Delta_{\min}^{-1} + 3\varepsilon^{-1})^2$ and $H_i(\mu, \varepsilon) = \Theta(K/\Delta_{i+1}^{-2})$. Ordered distinct mean gaps $(\Delta_i)_{i \in [C_\mu]}$ and $i_\mu(\tilde{\varepsilon}) = i$ if $\tilde{\varepsilon} \in [\Delta_i, \Delta_{i+1})$.

Policy playing $(\hat{\imath}_n)_{n > K}$:

☞ Anytime expected simple regret with exponential decay.

Proof sketch

❶ For all $\delta \in (0,1]$, let $T_{\tilde{\varepsilon}}(\delta)$ and $(\mathcal{E}_{n,\delta})_n$ such that $\max_n \mathbb{P}_\nu(\mathcal{E}_{n,\delta}^\complement) \le \delta$ and $\{\hat{\imath}_n \notin \mathcal{I}_{\tilde{\varepsilon}}(\mu)\} \subset \mathcal{E}_{n,\delta}^\complement$ for all $n > T_{\tilde{\varepsilon}}(\delta)$. Then,

$$\mathbb{P}_\nu(\hat{\imath}_n \notin \mathcal{I}_{\tilde{\varepsilon}}(\mu)) \le \inf\{\delta \mid \ n > T_{\tilde{\varepsilon}}(\delta)\} \, .$$

Proof sketch

❶ For all $\delta \in (0, 1]$, let $T_{\tilde{\varepsilon}}(\delta)$ and $(\mathcal{E}_{n,\delta})_n$ such that $\max_n \mathbb{P}_\nu(\mathcal{E}_{n,\delta}^\complement) \le \delta$ and $\{\hat{\imath}_n \notin \mathcal{I}_{\tilde{\varepsilon}}(\mu)\} \subset \mathcal{E}_{n,\delta}^\complement$ for all $n > T_{\tilde{\varepsilon}}(\delta)$. Then,

$$\mathbb{P}_\nu(\hat{\imath}_n \notin \mathcal{I}_{\tilde{\varepsilon}}(\mu)) \le \inf\{\delta \mid \ n > T_{\tilde{\varepsilon}}(\delta)\} \, .$$

❷ A necessary condition for error: undersampled arms still exist.

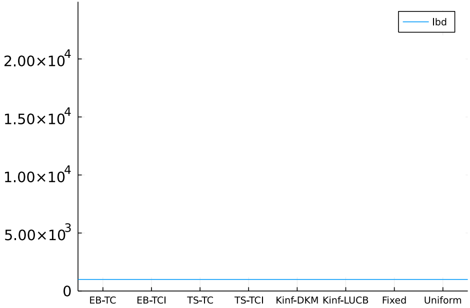
❸ If there are undersampled arms, there is an arm which is selected either as leader or challenger and has a bounded selection count.

Key observation: *The number of times one can increment a bounded positive variable by one is also bounded.*

Crop-management task

Bounded instance with $K = 4$ at $(\varepsilon, \delta) = (0, 10^{-2})$, Top Two with fixed design $\beta = 1/2$

arm = planting date / observation = bounded yield



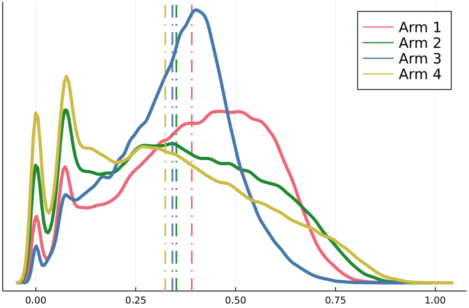


Figure: Empirical stopping time (a) on scaled DSSAT instances with their density and mean (b). Lower bound is $T_0(\nu)\log(1/\delta)$.

Crop-management task

Bounded instance with $K = 4$ at $(\varepsilon, \delta) = (0, 10^{-2})$, Top Two with fixed design $\beta = 1/2$

arm = planting date / observation = bounded yield

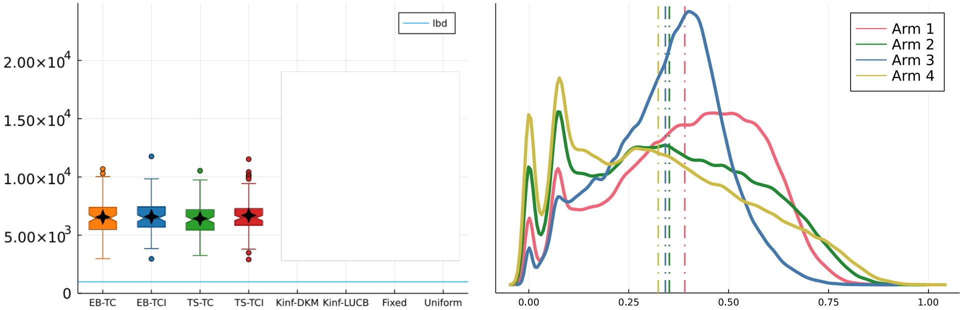
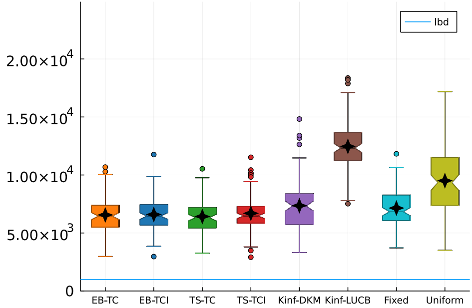


Figure: Empirical stopping time (a) on scaled DSSAT instances with their density and mean (b). Lower bound is $T_0(\nu) \log(1/\delta)$.

Bounded instance with $K = 4$ at $(\varepsilon, \delta) = (0, 10^{-2})$, Top Two with fixed design $\beta = 1/2$

arm = planting date / observation = bounded yield



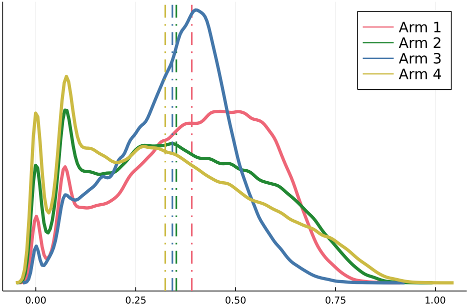
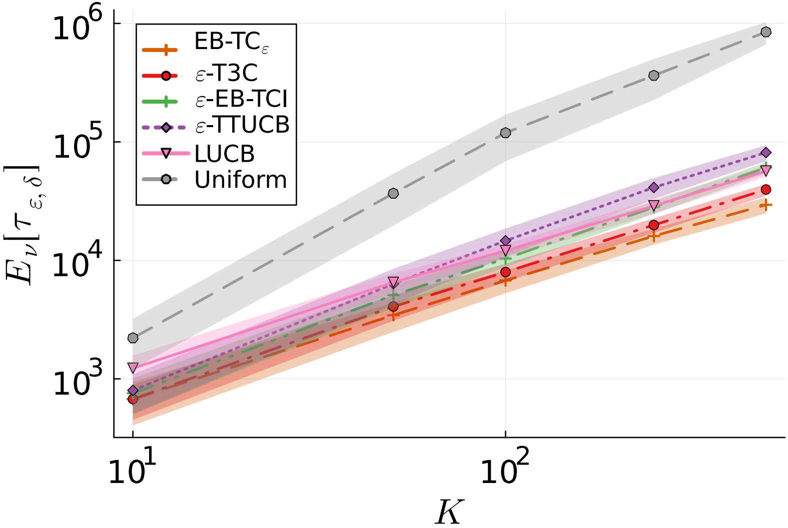


Figure: Empirical stopping time (a) on scaled DSSAT instances with their density and mean (b). Lower bound is $T_0(\nu) \log(1/\delta)$.

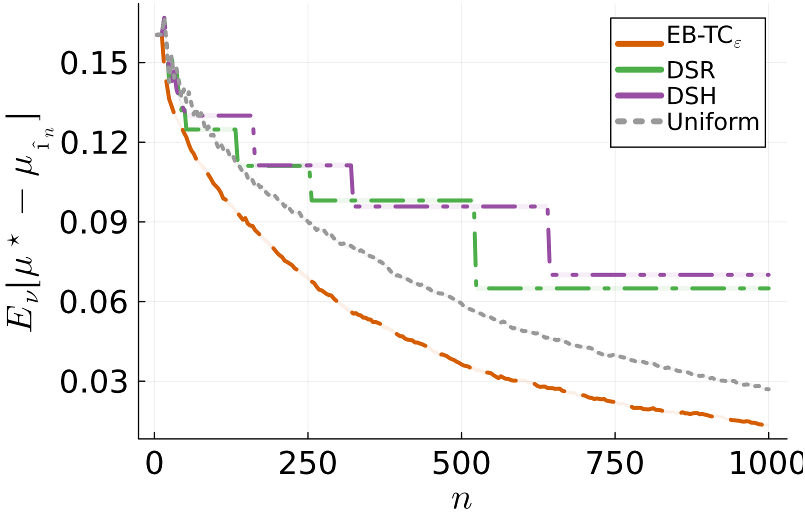
Empirical stopping time

Gaussian instances $\mu_i = 1 - \frac{(i-1)^\alpha}{(K-1)^\alpha}$ for $\alpha = 0.6$ with varying K at $(\varepsilon, \delta) = (10^{-1}, 10^{-2})$



Empirical simple regret

Gaussian instance $\mu \in \{0.6, 0.4\}^{10}$ with $|\mathcal{I}_0(\mu)| = 2$, EB-TC$_\varepsilon$ uses $(\varepsilon, \beta) = (0.1, 1/2)$



Transductive linear bandits

Mean vector $\theta \in \mathbb{R}^d$, set of arms $\mathcal{A} \subseteq \mathbb{R}^d$ and answers $\mathcal{Z} \subseteq \mathbb{R}^d$.

Goal: Identify one answer in $\mathcal{Z}_\varepsilon(\theta) = \{z \mid \langle \theta, z \rangle \geq \max_x \langle \theta, x \rangle - \varepsilon\}$.

Transductive linear bandits

Mean vector $\theta \in \mathbb{R}^d$, set of arms $\mathcal{A} \subseteq \mathbb{R}^d$ and answers $\mathcal{Z} \subseteq \mathbb{R}^d$.

Goal: Identify one answer in $\mathcal{Z}_\varepsilon(\theta) = \{z \mid \langle \theta, z \rangle \geq \max_x \langle \theta, x \rangle - \varepsilon\}$.

$$T_\varepsilon(\nu)^{-1} = \max_{z \in \mathcal{Z}_\varepsilon(\mu)} \max_{w \in \triangle_K} \min_{x \neq z} C_\varepsilon(z, x; \nu, w) \, .$$

Transductive linear bandits for Gaussian with unit variance

$$C_\varepsilon(z, x; \nu, w) = \mathbb{1}\left(\langle \theta, z - x \rangle > -\varepsilon\right) \frac{(\langle \theta, z - x \rangle + \varepsilon)^2}{2\|z - x\|^2_{V_w^{-1}}} \, ,$$

with $V_w = \sum_a w_a a a^\mathsf{T}$ is the design matrix of the allocation $w \in \triangle_K$.

The structured Top Two approach

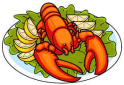
Set a **leader** answer $B_n \in \mathcal{Z}$;

Set a **challenger** answer $C_n \in \mathcal{Z} \setminus \{B_n\}$;

Set a **target** $\beta_n(B_n, C_n) \in \triangle_K$;

Return $I_n \in \mathcal{A}$ using target $\beta_n(B_n, C_n)$.

Answers



Arms

The LεTT algorithm

Subproblem: **known $\boldsymbol{\theta}$** and $\varepsilon = 0$, leader $z^\star = \arg\max_{z \in \mathcal{Z}} \langle \theta, z \rangle$.

The LεTT algorithm

Subproblem: **known $\boldsymbol{\theta}$** and $\varepsilon = 0$, leader $z^\star = \arg\max_{z \in \mathcal{Z}} \langle \theta, z \rangle$.

Sequentially learned components $(q_n, w_n) \in \triangle_{Z-1} \times \triangle_K$

☞ **TC challenger**, **Frank-Wolfe** step

$$C_n \in \arg\min_{x \neq z^\star} C(x, w_n) \quad \text{with} \quad C(x, w) = \frac{\langle \theta, z^\star - x \rangle^2}{2\|z^\star - x\|^2_{V_w^{-1}}} .$$

☞ **IDS target**, **normalized reweighted gradient** step

$$\beta_n(C_n) = w_n \odot \nabla_w C(C_n, w_n) / C(C_n, w_n) .$$

Then, update $\begin{bmatrix} q_{n+1} \\ w_{n+1} \end{bmatrix} = \left(1 - \frac{1}{n+1}\right) \begin{bmatrix} q_n \\ w_n \end{bmatrix} + \frac{1}{n+1} \begin{bmatrix} \mathbf{1}_{C_n} \\ \boldsymbol{\beta_n(C_n)} \end{bmatrix}$.

The LεTT algorithm

Subproblem: **known $\boldsymbol{\theta}$** and $\varepsilon = 0$, leader $z^\star = \arg\max_{z \in \mathcal{Z}} \langle \theta, z \rangle$.

Sequentially learned components $(q_n, w_n) \in \triangle_{Z-1} \times \triangle_K$

☞ **TC challenger**, **Frank-Wolfe** step

$$C_n \in \arg\min_{x \neq z^\star} C(x, w_n) \quad \text{with} \quad C(x, w) = \frac{\langle \theta, z^\star - x \rangle^2}{2 \| z^\star - x \|^2_{V_w^{-1}}} .$$

☞ **IDS target**, **normalized reweighted gradient** step

$$\beta_n(C_n) = w_n \odot \nabla_w C(C_n, w_n) / C(C_n, w_n) .$$

Then, update $\begin{bmatrix} q_{n+1} \\ w_{n+1} \end{bmatrix} = (1 - \frac{1}{n+1}) \begin{bmatrix} q_n \\ w_n \end{bmatrix} + \frac{1}{n+1} \begin{bmatrix} \mathbf{1}_{C_n} \\ \boldsymbol{\beta}_n(C_n) \end{bmatrix}$.

Open problem: **Show the convergence towards a saddle point of**

$$\max_{w \in \triangle_K} \min_{q \in \triangle_{Z-1}} \langle q, C(\cdot, w) \rangle .$$

Conclusion

The **Top Two approach** meets our requirements !

To be advocated by statisticians:

✓ guarantees on the quality of the recommendation,

✓ empirically competitive.

To be used by practitioners:

✓ simple,

✓ interpretable,

✓ generalizable,

✓ versatile.

The **Top Two approach** meets our requirements !

To be advocated by statisticians:

✓ guarantees on the quality of the recommendation,

✓ empirically competitive.

To be used by practitioners:

✓ simple,

✓ interpretable,

✓ generalizable,

✓ versatile.

Perspectives:

- structured Top Two approach,
- anytime setting,
- privacy, safety and fairness.

References I

Agrawal, S., Juneja, S., and Glynn, P. W. (2020). Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory (ALT)*.

Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best Arm Identification in Multi-armed Bandits. In *Conference on Learning Theory*.

Carpentier, A. and Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proceedings of the 29th Conference on Learning Theory (COLT)*.

Degenne, R. (2023). On the existence of a complexity in fixed budget bandit identification. *International Conference on Learning Theory (COLT)*.

Degenne, R. and Koolen, W. M. (2019). Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Degenne, R., Koolen, W. M., and Ménard, P. (2019). Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Degenne, R., Shao, H., and Koolen, W. M. (2020). Structure adaptive algorithms for stochastic bandits. In *International Conference on Machine Learning (ICML)*.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference On Learning Theory*.

Jourdan, M. and Degenne, R. (2023). Non-asymptotic analysis of a ucb-based top two algorithm. *Thirty-Seventh Conference on Neural Information Processing Systems*.

Jourdan, M., Degenne, R., Baudry, D., De Heide, R., and Kaufmann, E. (2022). Top two algorithms revisited. *Advances in Neural Information Processing Systems*.

References II

Jourdan, M., Degenne, R., and Kaufmann, E. (2023a). Dealing with unknown variances in best-arm identification. *International Conference on Algorithmic Learning Theory*.

Jourdan, M., Degenne, R., and Kaufmann, E. (2023b). An ε-best-arm identification algorithm for fixed-confidence and beyond. *Thirty-Seventh Conference on Neural Information Processing Systems*.

Jourdan, M. and Réda, C. (2023). An anytime algorithm for good arm identification. *arXiv preprint arXiv:2310.10359*.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. (2012). PAC subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning (ICML)*.

Kano, H., Honda, J., Sakamaki, K., Matsuura, K., Nakamura, A., and Sugiyama, M. (2019). Good arm identification via bandit feedback. *Machine Learning*, 108(5):721–745.

Karnin, Z., Koren, T., and Somekh, O. (2013). Almost optimal Exploration in multi-armed bandits. In *International Conference on Machine Learning (ICML)*.

Komiyama, J., Tsuchiya, T., and Honda, J. (2022). Minimax optimal algorithms for fixed-budget best arm identification. In *Advances in Neural Information Processing Systems*.

Qin, C., Klabjan, D., and Russo, D. (2017). Improving the expected improvement algorithm. In *Advances in Neural Information Processing Systems 30 (NIPS)*.

Riou, C. and Honda, J. (2020). Bandit algorithms based on thompson sampling for bounded reward distributions. In *Algorithmic Learning Theory (ALT)*.

Russo, D. (2016). Simple Bayesian algorithms for best arm identification. In *Proceedings of the 29th Conference on Learning Theory (COLT)*.

References III

Shang, X., de Heide, R., Kaufmann, E., Ménard, P., and Valko, M. (2020). Fixed-confidence guarantees for bayesian best-arm identification. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Wang, P.-A., Ariu, K., and Proutiere, A. (2024). On universally optimal algorithms for a/b testing.

Wang, P.-A., Tzeng, R.-C., and Proutiere, A. (2021). Fast pure exploration via frank-wolfe. *Advances in Neural Information Processing Systems*.

You, W., Qin, C., Wang, Z., and Yang, S. (2023). Information-directed selection for top-two algorithms. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2850–2851. PMLR.