# Top Two Algorithms Revisited

Marc Jourdan, Rémy Degenne, Dorian Baudry,
Rianne de Heide and Émilie Kaufmann

June 17, 2022

Centre de Recherche en Informatique, Signal et Automatique de Lille
Université de Lille
VRIJE UNIVERSITEIT AMSTERDAM

Section 1

# Motivation

# Motivation

**Goal:** Identify the item having the highest averaged return.

Applications:

- A/B testing for online marketing,
- phase II/III of clinical trials,
- crop-management tasks.

# Statistical model

Frequent distributions: parametric, e.g. Bernoulli or Gaussian.

Applications:

- ✓ A/B testing for online marketing,
- ✓ phase II/III of clinical trials,
- ✗ crop-management tasks.

*Nature is bounded:*

☞ Bounded distributions

# Statistical model

Frequent distributions: parametric, e.g. Bernoulli or Gaussian.

Applications:

- ✓ A/B testing for online marketing,
- ✓ phase II/III of clinical trials,
- ✗ crop-management tasks.

*Nature is bounded:*

☞ Bounded distributions

# Crop-management

Simulator of crop yield:

- $30$ years of historical field data for $42$ different plants and soil conditions,
- model complex biophysical processes.

Case study:

- maize fields with Sub-Saharan soil conditions,
- fixed fertilization policy,
- identify the best planting date.

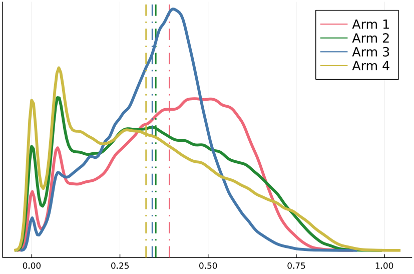


Figure: Decision Support System for Agrotechnology Transfer

# Crop-management

Simulator of crop yield:

- $30$ years of historical field data for $42$ different plants and soil conditions,
- model complex biophysical processes.

Case study:

- maize fields with Sub-Saharan soil conditions,
- fixed fertilization policy,
- identify the best planting date.

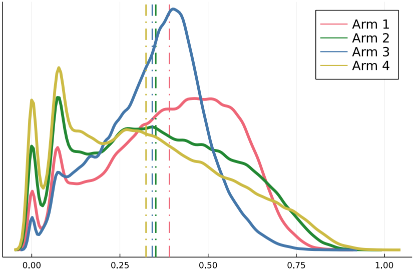


Figure: Decision Support System for Agrotechnology Transfer

# Section 2

## Problem statement

# Stochastic multi-armed bandits

$K$ arms, $F_i$ cdf of arm $i$ with mean $m(F_i) := \mathbb{E}_{X \sim F_i}[X]$.

At time $n$, pull $I_n \in [K]$ and observe $X_{n,I_n} \sim F_{I_n}$.

Distributions $\mathcal{F}$ with set of possible means $\mathcal{I}$:

- bounded in $[0, B]$,
- sub-exponential single parameter exponential families (SPEF, e.g. Bernoulli, Gaussian with known variance, etc).

# Stochastic multi-armed bandits

$K$ arms, $F_i$ cdf of arm $i$ with mean $m(F_i) := \mathbb{E}_{X \sim F_i}[X]$.

At time $n$, pull $I_n \in [K]$ and observe $X_{n,I_n} \sim F_{I_n}$.

Distributions $\mathcal{F}$ with set of possible means $\mathcal{I}$:

- bounded in $[0, B]$,
- sub-exponential single parameter exponential families (SPEF, e.g. Bernoulli, Gaussian with known variance, etc).

# Best-arm identification (BAI)

**Goal:** identify the best arm $i^\star = \arg\max_i m(F_i)$ with confidence $\delta$.

- sampling rule, $I_n \in [K]$,
- recommendation rule, $\hat{\imath}_n \in [K]$,
- stopping rule, $\tau_\delta$.

**Objective:** Minimize $\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]$ for $\delta$-correct algorithms

$$\mathbb{P}_{\boldsymbol{F}}[\tau_\delta < +\infty, \, \hat{\imath}_{\tau_\delta} \neq i^\star] \leq \delta \, .$$

? What is the best one could achieve ? Agrawal et al. (2020)

☞ For all $\delta$-correct algorithm, for all $\boldsymbol{F} \in \mathcal{F}^K$,

$$\mathbb{E}_{\boldsymbol{F}}[\tau_\delta] \geq T^\star(\boldsymbol{F}) \log\left(1/(2.4\delta)\right) \, .$$

# Best-arm identification (BAI)

**Goal:** identify the best arm $i^\star = \arg\max_i m(F_i)$ with confidence $\delta$.

- sampling rule, $I_n \in [K]$,
- recommendation rule, $\hat{\imath}_n \in [K]$,
- stopping rule, $\tau_\delta$.

**Objective:** Minimize $\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]$ for $\delta$-correct algorithms

$$\mathbb{P}_{\boldsymbol{F}}[\tau_\delta < +\infty,\ \hat{\imath}_{\tau_\delta} \neq i^\star] \leq \delta \,.$$

? What is the best one could achieve ? Agrawal et al. (2020)

☞ For all $\delta$-correct algorithm, for all $\boldsymbol{F} \in \mathcal{F}^K$,

$$\mathbb{E}_{\boldsymbol{F}}[\tau_\delta] \geq T^\star(\boldsymbol{F}) \log\left(1/(2.4\delta)\right) \,.$$

# Characteristic time

$$T^\star(\boldsymbol{F})^{-1} := \sup_{w \in \triangle_K} \min_{i \neq i^\star} \inf_{u \in \mathcal{I}} \left\{ w_{i^\star} \mathcal{K}_{\inf}^{-}(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^{+}(F_i, u) \right\} \ ,$$

$\triangle_K$ simplex, $\mathcal{K}_{\inf}^{\pm}(F, u) := \inf \left\{ \mathrm{KL}(F, G) \mid G \in \mathcal{F}, \ m(G) \gtrless u \right\}$.

? How can we reach the lower bound $T^\star(\boldsymbol{F})$ ?

☞ Empirical sampling proportions converging towards maximizer.

**Problem:** learning the maximizer $w^\star \in \triangle_K$ can be difficult.

Observation: the allocation to the best arm has a central role.

# Characteristic time

$$T^{\star}(\boldsymbol{F})^{-1} := \sup_{w \in \triangle_K} \min_{i \neq i^{\star}} \inf_{u \in \mathcal{I}} \left\{ w_{i^{\star}} \mathcal{K}_{\text{inf}}^{-}(F_{i^{\star}}, u) + w_i \mathcal{K}_{\text{inf}}^{+}(F_i, u) \right\} ,$$

$\triangle_K$ simplex, $\mathcal{K}_{\text{inf}}^{\pm}(F, u) := \inf \left\{ \text{KL}(F, G) \mid G \in \mathcal{F},\ m(G) \gtrless u \right\}$.

? How can we reach the lower bound $T^{\star}(\boldsymbol{F})$ ?

☞ Empirical sampling proportions converging towards maximizer.

**Problem:** learning the maximizer $w^{\star} \in \triangle_K$ can be difficult.

Observation: the allocation to the best arm has a central role.

# Asymptotic $\beta$-optimality

Sub-class of algorithms: $\beta$ proportion of samples to the best arm.

? What is the best one could achieve ? Russo (2016)

☞ $\beta$-optimal: $\limsup_{\delta \to 0} \mathbb{E}_{\boldsymbol{F}}[\tau_\delta] / \log(1/\delta) \leq T^\star_\beta(\boldsymbol{F})$ where

$$T^\star_\beta(\boldsymbol{F})^{-1} := \sup_{w \in \triangle_K, w_{i^\star} = \beta} \min_{i \neq i^\star} \inf_{u \in \mathcal{I}} \left\{ \beta \mathcal{K}^-_{\text{inf}}(F_{i^\star}, u) + w_i \mathcal{K}^+_{\text{inf}}(F_i, u) \right\} ,$$

achieved for a unique $\beta$-optimal allocation $w^\beta$ when $i^\star$ is unique.

? How does it relate to asymptotic optimality ?

☞ $T^\star(\boldsymbol{F}) = \min_{\beta \in (0,1)} T^\star_\beta(\boldsymbol{F})$ and $T^\star_{1/2}(\boldsymbol{F}) \leq 2T^\star(\boldsymbol{F})$.

# Asymptotic $\beta$-optimality

Sub-class of algorithms: $\beta$ proportion of samples to the best arm.

? What is the best one could achieve ? Russo (2016)

☞ $\beta$-optimal: $\limsup_{\delta \to 0} \mathbb{E}_{\boldsymbol{F}}[\tau_\delta]/\log(1/\delta) \leq T^\star_\beta(\boldsymbol{F})$ where

$$T^\star_\beta(\boldsymbol{F})^{-1} := \sup_{w \in \triangle_K, w_{i^\star} = \beta} \min_{i \neq i^\star} \inf_{u \in \mathcal{I}} \left\{ \beta \mathcal{K}^-_{\inf}(F_{i^\star}, u) + w_i \mathcal{K}^+_{\inf}(F_i, u) \right\} ,$$

achieved for a unique $\beta$-optimal allocation $w^\beta$ when $i^\star$ is unique.

? How does it relate to asymptotic optimality ?

☞ $T^\star(\boldsymbol{F}) = \min_{\beta \in (0,1)} T^\star_\beta(\boldsymbol{F})$ and $T^\star_{1/2}(\boldsymbol{F}) \leq 2T^\star(\boldsymbol{F})$.

# Contributions

1. Generic and modular analysis of Top Two algorithms.
2. Asymptotically $\beta$-optimal instances (bounded and SPEF).
3. Competitive performance on a real-world non-parametric task.

# Related work

Top Two (TT) algorithms for Gaussians:

- Russo (2016), TTPS and TTTS (Probability/Thompson Sampling),
- Qin et al. (2017), TTEI (Expected Improvement),
- Shang et al. (2020), T3C (Transportation Cost).

Other BAI algorithms:

- Kalyanakrishnan et al. (2012), (kl)-LUCB algorithm for bounded distributions,
- Agrawal et al. (2020), Track-and-Stop for heavy-tailed distributions,
- Degenne et al. (2019), DKM for sub-Gaussian SPEF.

# Related work

Top Two (TT) algorithms for Gaussians:

- Russo (2016), TTPS and TTTS (Probability/Thompson Sampling),
- Qin et al. (2017), TTEI (Expected Improvement),
- Shang et al. (2020), T3C (Transportation Cost).

Other BAI algorithms:

- Kalyanakrishnan et al. (2012), (kl)-LUCB algorithm for bounded distributions,
- Agrawal et al. (2020), Track-and-Stop for heavy-tailed distributions,
- Degenne et al. (2019), DKM for sub-Gaussian SPEF.

# Section 3

## Top Two algorithms

# Stopping-recommendation pair

? Which arm should we recommend ?

$$\hat{\imath}_n = \arg\max_i \mu_{n,i} \quad \text{with} \quad \mu_{n,i} = m(F_{n,i}) \ ,$$

$N_{n,i} = \sum_{t\in[n]} \mathbb{1}\left(I_t = i\right)$ and $F_{n,i} = \frac{1}{N_{n,i}} \sum_{t\in[n]} \delta_{X_{t,I_t}} \mathbb{1}\left(I_t = i\right)$.

? How to stop to obtain $\delta$-correct algorithm ?

☞ calibrated **GLR stopping rule**

$$\tau_\delta = \inf\left\{ n \in \mathbb{N} \mid \min_{j\neq\hat{\imath}_n} W_n(\hat{\imath}_n, j) > \beta(n,\delta) \right\} \ , \tag{1}$$

where the empirical transportation cost between arms $(i,j)$ is

$$W_n(i,j) = \inf_{x\in\mathcal{I}} \left[ N_{n,i}\mathcal{K}_{\inf}^{-}(F_{n,i}, x) + N_{n,j}\mathcal{K}_{\inf}^{+}(F_{n,j}, x) \right] \ .$$

# Stopping-recommendation pair

❓ Which arm should we recommend ?

$$\hat{\imath}_n = \arg\max_i \mu_{n,i} \quad \text{with} \quad \mu_{n,i} = m(F_{n,i}) \,,$$

$N_{n,i} = \sum_{t \in [n]} \mathbb{1}\left(I_t = i\right)$ and $F_{n,i} = \frac{1}{N_{n,i}} \sum_{t \in [n]} \delta_{X_{t,I_t}} \mathbb{1}\left(I_t = i\right)$.

❓ How to stop to obtain $\delta$-correct algorithm ?

☞ calibrated **GLR stopping rule**

$$\tau_\delta = \inf \left\{ n \in \mathbb{N} \mid \min_{j \neq \hat{\imath}_n} W_n(\hat{\imath}_n, j) > \beta(n, \delta) \right\} \,, \tag{1}$$

where the empirical transportation cost between arms $(i, j)$ is

$$W_n(i,j) = \inf_{x \in \mathcal{I}} \left[ N_{n,i} \mathcal{K}^-_{\inf}(F_{n,i}, x) + N_{n,j} \mathcal{K}^+_{\inf}(F_{n,j}, x) \right] \,.$$

# Sampling rule

? How should we pull arms with the $\beta$ constraint ?

☞ **Top Two sampling rule** with fixed $\beta$ !

1: Choose a **leader** $B_n \in [K]$
2: $U \sim \mathcal{U}([0, 1])$
3: **if** $U < \beta$ **then**
4: $\quad I_n = B_n$
5: **else**
6: $\quad$ Choose a **challenger** $C_n \in [K] \setminus \{B_n\}$
7: $\quad I_n = C_n$
8: **end if**
9: **Output**: next arm to sample $I_n$

# Choices of leader

Empirical Best (**EB**), deterministic,

$$B_n^{\mathsf{EB}} \in \arg\max_{i \in [K]} \mu_{n-1,i} \ .$$

Thompson Sampling (**TS**), randomized with a sampler $\Pi_n$ on $\mathcal{I}^K$,

$$B_n^{\mathsf{TS}} \in \arg\max_{i \in [K]} \theta_i \quad \text{where} \quad \theta \sim \Pi_{n-1} \ .$$

# Choices of challenger given leader $B_n$

Transportation Cost (**TC**), deterministic,

$$C_n^{\mathsf{TC}} \in \arg\min_{j \neq B_n} W_{n-1}(B_n, j) \; .$$

Transportation Cost Improved (**TCI**), deterministic,

$$C_n^{\mathsf{TCI}} \in \arg\min_{j \neq B_n} W_{n-1}(B_n, j) + \log N_{n-1,j} \; .$$

Re-Sampling (**RS**), randomized, repeat $\theta \sim \Pi_{n-1}$ until

$$C_n^{\mathsf{RS}} \in \arg\max_{i \in [K]} \theta_i \not\ni B_n \; .$$

# Novelties

Six instances denoted by $\beta$-[leader]-[challenger].

Literature:

- TTTS and T3C corresponds to $\beta$-TS-RS and $\beta$-TS-TC.
- $\beta$-optimality for Gaussian distributions.

Novelties:

- Fully deterministic instances are possible with the EB leader.
- The TCI challenger is more stable than the TC one by penalizing over-sampled challengers.
- Dirichlet sampler for BAI with bounded distributions.
- Bounded distributions and SPEF of sub-exponential distributions.

# Novelties

Six instances denoted by $\beta$-[leader]-[challenger].

Literature:

- TTTS and T3C corresponds to $\beta$-TS-RS and $\beta$-TS-TC.
- $\beta$-optimality for Gaussian distributions.

Novelties:

- Fully deterministic instances are possible with the EB leader.
- The TCI challenger is more stable than the TC one by penalizing over-sampled challengers.
- Dirichlet sampler for BAI with bounded distributions.
- Bounded distributions and SPEF of sub-exponential distributions.

Section 4

# Bounded distributions

# Deterministic instances

Threshold ensuring $\delta$-correctness of the stopping rule (1)

$$\beta(n,\delta) = \log(1/\delta) + 2\log(1+n/2) + 2 + \log(K-1)\,. \tag{2}$$

Computing transportation costs between arm $i$ and arm $j$:

$$N_{n,i}\mathcal{K}_{\inf}^{+}(F_{n,i},x) = \sup_{\lambda\in[0,1]} \sum_{t\in[N_{n,i}]} \log\left(1-\lambda\frac{X_{t,i}-x}{B-x}\right)\,.$$

# Randomized instances

? How to design a sampler over $(0, B)^K$ ? Riou and Honda (2020)

☞ **Dirichlet sampler:** $\Pi_n = \bigtimes_{i \in [K]} \Pi_{n,i}$ where $\Pi_{n,i}$ uses the empirical cdf $F_{n,i}$ augmented by $\{0, B\}$. The sampler $\Pi_{n,i}$ returns

$$\sum_{t \in [N_{n,i}]} w_t X_{t,i} + B w_{N_{n,i}+1} \quad \text{with} \quad \boldsymbol{w} \sim \mathrm{Dir}(\mathbf{1}_{N_{n,i}+2}) \, .$$

# Sample complexity upper bound

## Theorem

*Combining the stopping rule (1) with threshold (2) and a Top Two algorithm with $\beta \in (0,1)$, instantiated with any pair of leader/challenger introduced above, yields a $\delta$-correct algorithm which is asymptotically $\beta$-optimal for all $\boldsymbol{F} \in \mathcal{F}^K$ with $m(\boldsymbol{F}) \in (0,B)^K$ and $\Delta_{\min} := \min_{i \neq j} |m(F_i) - m(F_j)| > 0$.*

Distinct means:

- Uncommon, used for sufficient exploration of Top Two algorithms.
- Good empirical performance even when $\Delta_{\min} = 0$.

# Sample complexity upper bound

## Theorem

*Combining the stopping rule (1) with threshold (2) and a Top Two algorithm with $\beta \in (0, 1)$, instantiated with any pair of leader/challenger introduced above, yields a $\delta$-correct algorithm which is asymptotically $\beta$-optimal for all $\boldsymbol{F} \in \mathcal{F}^K$ with $m(\boldsymbol{F}) \in (0, B)^K$ and $\Delta_{\min} := \min_{i \neq j} |m(F_i) - m(F_j)| > 0$.*

Distinct means:

- Uncommon, used for sufficient exploration of Top Two algorithms.
- Good empirical performance even when $\Delta_{\min} = 0$.

# Comparing instances

**Limitations:**

- For large $n$, the RS challenger is computationally costly and the TS leader is expensive.
- $\beta$-EB-TC is too greedy and lacks robustness for moderate regime.

**Advantages:**

- The EB leader is computationally efficient and the TC(I) challengers are not costlier than computing the stopping rule.
- The TS leader and the TCI challenger foster implicit exploration.

**Recommendations:** $\beta$-EB-TCI and $\beta$-TS-TC.

# Comparing instances

**Limitations:**

- For large $n$, the RS challenger is computationally costly and the TS leader is expensive.
- $\beta$-EB-TC is too greedy and lacks robustness for moderate regime.

**Advantages:**

- The EB leader is computationally efficient and the TC(I) challengers are not costlier than computing the stopping rule.
- The TS leader and the TCI challenger foster implicit exploration.

**Recommendations:** $\beta$-EB-TCI and $\beta$-TS-TC.

# Section 5

## Modular sample complexity analysis

# Reaching asymptotic $\beta$-optimality

**?** How can we reach asymptotic $\beta$-optimality ?

☞ Empirical proportions converging towards maximizer $w^{\beta}$.

**Convergence time** $T_{\beta}^{\varepsilon}$ defined as

$$T_{\beta}^{\varepsilon} := \inf \left\{ T \geq 1 \mid \forall n \geq T,\ \max_{i \in [K]} \left| \frac{N_{n,i}}{n} - w_i^{\beta} \right| \leq \varepsilon \right\} .$$

For any sampling rule, there exists $\varepsilon_0(\boldsymbol{F}) > 0$,

$$\forall \varepsilon \in (0, \varepsilon_0(\boldsymbol{F})],\ \mathbb{E}_{\boldsymbol{F}}[T_{\beta}^{\varepsilon}] < +\infty \quad \Longrightarrow \quad \limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_{\delta}]}{\log(1/\delta)} \leq T_{\beta}^{\star}(\boldsymbol{F}) .$$

# Reaching asymptotic $\beta$-optimality

**?** How can we reach asymptotic $\beta$-optimality ?

☞ Empirical proportions converging towards maximizer $w^\beta$.

**Convergence time** $T^\varepsilon_\beta$ defined as

$$T^\varepsilon_\beta := \inf \left\{ T \geq 1 \mid \forall n \geq T, \max_{i \in [K]} \left| \frac{N_{n,i}}{n} - w^\beta_i \right| \leq \varepsilon \right\} .$$

For any sampling rule, there exists $\varepsilon_0(\boldsymbol{F}) > 0$,

$$\forall \varepsilon \in (0, \varepsilon_0(\boldsymbol{F})], \ \mathbb{E}_{\boldsymbol{F}}[T^\varepsilon_\beta] < +\infty \quad \implies \quad \limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T^\star_\beta(\boldsymbol{F}) .$$

# Proving $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$

❶ **Sufficient exploration**: for $n$ large enough,

$$\min_{i \in [K]} N_{n,i} \geq \sqrt{n/K} \, .$$

☞ Combining properties of the leader and challenger.

❷ Convergence of $\frac{N_n}{n}$ towards $w^\beta$ under sufficient exploration.

Let $\psi_{n,i} := \mathbb{P}_{|(n-1)}[I_n = i]$ and $\Psi_{n,i} := \sum_{t \in [n]} \psi_{t,i}$,

☞ $(N_{n,i} - \Psi_{n,i})/\sqrt{n}$ are sub-Gaussian random variables.

# Proving $\mathbb{E}_{\boldsymbol{F}}[T_{\beta}^{\varepsilon}] < +\infty$

❶ **Sufficient exploration**: for $n$ large enough,

$$\min_{i \in [K]} N_{n,i} \geq \sqrt{n/K} \, .$$

☞ Combining properties of the leader and challenger.

❷ Convergence of $\frac{N_n}{n}$ towards $w^{\beta}$ under sufficient exploration.

Let $\psi_{n,i} := \mathbb{P}_{|(n-1)}[I_n = i]$ and $\Psi_{n,i} := \sum_{t \in [n]} \psi_{t,i}$,

☞ $(N_{n,i} - \Psi_{n,i})/\sqrt{n}$ are sub-Gaussian random variables.

# Convergence towards $w^\beta$: leader's property

$$\psi_{n,i} = \beta \mathbb{P}_{|(n-1)}[B_n = i] + (1-\beta) \sum_{j \neq i} \mathbb{P}_{|(n-1)}[B_n = j] \mathbb{P}_{|(n-1)}[C_n = i | B_n = j] \, .$$

For all $M \in \mathbb{N}$,

$$\left| \frac{\Psi_{n,i^\star}}{n} - \beta \right| \leq \frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] \, .$$

**Good leader:** for $n$ large enough,

$$\mathbb{P}_{|n}[B_{n+1} \neq i^\star] \leq g(n) =_{+\infty} o(n^{-\alpha}) \, .$$

# Convergence towards $w^\beta$: challenger's property

$$\psi_{n,i} = \beta \mathbb{P}_{|(n-1)}[B_n = i] + (1-\beta) \sum_{j \neq i} \mathbb{P}_{|(n-1)}[B_n = j] \mathbb{P}_{|(n-1)}[C_n = i | B_n = j] \, .$$

For all $M \in \mathbb{N}$ and all $i \neq i^\star$,

$$\frac{\Psi_{n,i}}{n} \leq \frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star] \, .$$

**Good challenger:** for $n$ large enough and all $i \neq i^\star$,

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \, \Rightarrow \, \mathbb{P}_{|n}[C_{n+1} = i | B_{n+1} = i^\star] \leq h(n) =_{+\infty} o(n^{-\alpha}) \, ,$$

## Section 6

## Experiments

# Experimental setup

Moderate regime, $\delta = 0.01$. Top Two algorithms with $\beta = \frac{1}{2}$.

Examples:

- Real-world non-parametric crop-management task,
- Random Bernoulli instances.

Benchmarks:

- KL-LUCB, "fixed" oracle and uniform sampling.
- Heuristics: $\mathcal{K}_{\text{inf}}$-DKM and $\mathcal{K}_{\text{inf}}$-LUCB.

# Crop-management problem

DSSAT: yield (observation) depending on the planting date (arm).

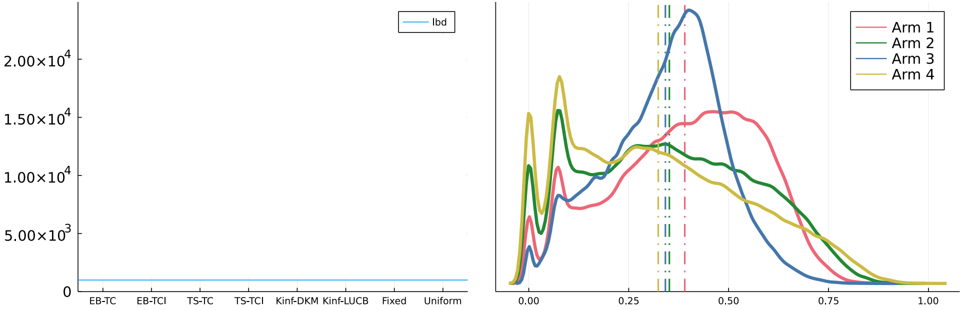


Figure: Empirical stopping time (a) on scaled DSSAT instances with their density and mean (b). Lower bound is $T^\star(\boldsymbol{F}) \log(1/\delta)$.

# Crop-management problem

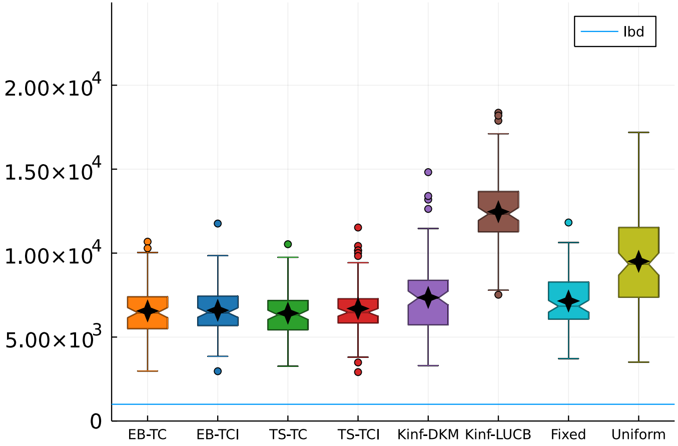


Figure: Empirical stopping time on scaled DSSAT instances. Lower bound is $T^\star(\boldsymbol{F})\log(1/\delta)$. “stars” equal means.

# Random Bernoulli instances

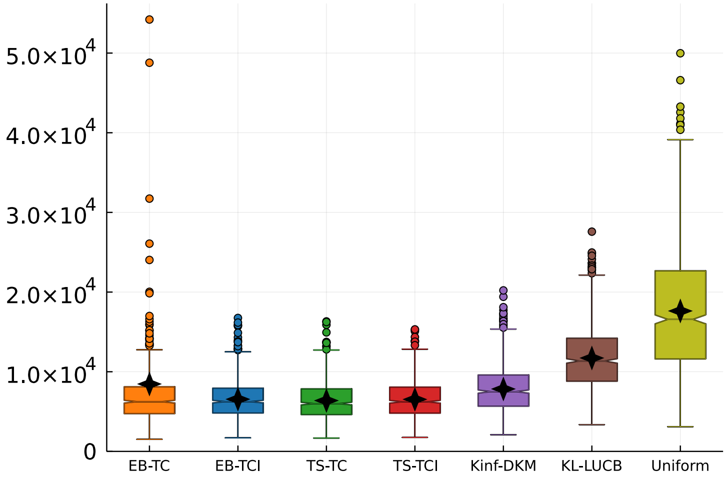


Figure: Empirical stopping time on random Bernoulli instances with $K = 10$.

# Conclusion

Contributions:

1. Generic and modular analysis of Top Two algorithms.
2. Asymptotically $\beta$-optimal instances (bounded and SPEF).
3. Competitive performance on a real-world non-parametric task.

Future work and open problems:

- Adaptive Top Two algorithms.
- Guarantees when $\Delta_{\min} = 0$.
- Fixed-budget setting.

# Conclusion

Contributions:

1. Generic and modular analysis of Top Two algorithms.
2. Asymptotically $\beta$-optimal instances (bounded and SPEF).
3. Competitive performance on a real-world non-parametric task.

Future work and open problems:

- Adaptive Top Two algorithms.
- Guarantees when $\Delta_{\min} = 0$.
- Fixed-budget setting.

# References


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

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

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)*.

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)*.

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)*.

Questions ?

# Appendix

# Distinct means

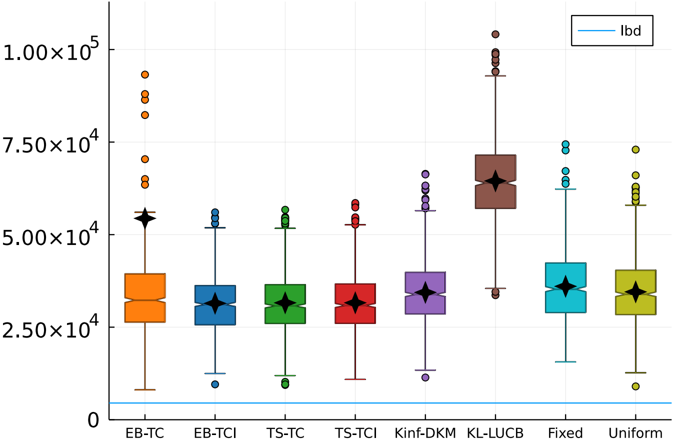


Figure: Empirical stopping time on Bernoulli instance $\mu = (0.5, 0.45, 0.45)$.

# RS challenger

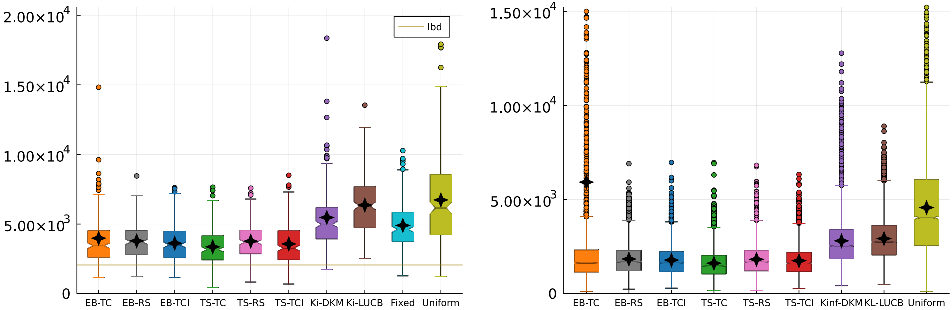


Figure: Empirical stopping time on (a) scaled DSSAT instances with $K = 6$ and (b) random Bernoulli instances with $K = 10$.

# $\mathcal{K}_{\inf}$-LUCB

$$\forall i \neq \hat{\imath}_n, \quad U_{n+1,i} = \max\left\{u \in [\mu_{n,i}, B] \mid N_{n,i}\mathcal{K}_{\inf}^{+}(F_{n,i}, u) \leq \beta(n,\delta)\right\} \;,$$
$$L_{n+1,\hat{\imath}_n} = \min\left\{u \in [0, \mu_{n,\hat{\imath}_n}] \mid N_{n,\hat{\imath}_n}\mathcal{K}_{\inf}^{-}(F_{n,\hat{\imath}_n}, u) \leq \beta(n,\delta)\right\} \;.$$

**Sampling rule:** Sample $B_n = \hat{\imath}_n$ *and* $C_n \in \arg\max_{i \neq \hat{\imath}_n} U_{n+1,i}$

**Stopping rule:**

$$\tau_\delta = \inf\left\{n \in \mathbb{N} \mid L_{n+1,\hat{\imath}_n} \geq \max_{j \neq \hat{\imath}_n} U_{n+1,j}\right\} \;.$$

# Convergence implies optimality

$$\mathbb{E}_{\boldsymbol{F}}[T_{\beta}^{\varepsilon}] < +\infty \quad \implies \quad \limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T_{\beta}^{\star}(\boldsymbol{F}) \ ,$$

Up to technicalities ($\mathcal{K}_{\inf}$ continuity and second order terms), this implication is shown by using that if $\tau_\delta \geq n$, then

$$\log(1/\delta) \approx_{\delta \to 0} \beta(n, \delta) \geq \min_{j \neq \hat{\imath}_n} W_n(\hat{\imath}_n, j) \approx_{n \geq T_{\beta}^{\varepsilon}} n T_{\beta}^{\star}(\boldsymbol{F})^{-1} \ .$$

It holds for bounded distributions and SPEF of sub-exponential distributions.

# Drawings