Top Two Algorithms Revisited

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

August 23, 2022

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.

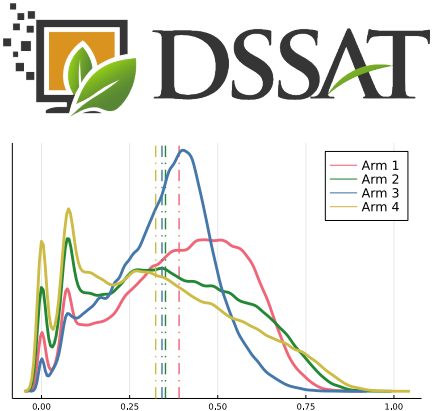

Figure: Decision Support System for Agrotechnology Transfer

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.

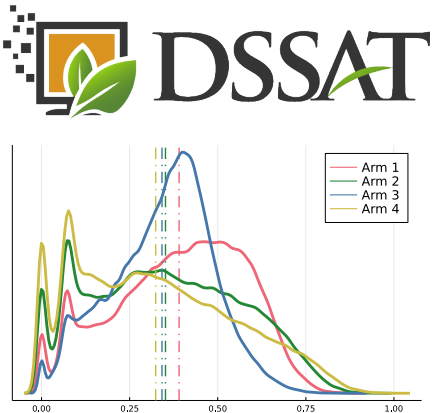

Figure: Decision Support System for Agrotechnology Transfer

Section 2

Problem Statement

Stochastic multi-armed bandits

K arms, $F_i \in \mathcal{F}$ 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 $\mathcal{F}$ are bounded on $[0, B]$.

Best-arm identification (BAI) in the fixed-confidence setting:
☞ identify $i^\star = \arg\max_i m(F_i)$ with confidence δ.

Identification strategy:

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

Stochastic multi-armed bandits

K arms, $F_i \in \mathcal{F}$ 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 $\mathcal{F}$ are bounded on $[0, B]$.

Best-arm identification (BAI) in the fixed-confidence setting:
☞ identify $i^\star = \arg\max_i m(F_i)$ with confidence δ.

Identification strategy:

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

Characteristic time

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

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

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

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

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

$$T^\star(\boldsymbol{F})^{-1} := \sup_{w \in \triangle_K} \inf_{\boldsymbol{G} \in \mathcal{F}^K :\, i^\star \notin i^\star(\boldsymbol{G})} \sum_{i \in [K]} w_i \mathrm{KL}(F_i, G_i)$$

$$= \sup_{w \in \triangle_K} \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\}\ ,$$

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

Asymptotic β-optimality

Sub-class of algorithms: β proportion of samples to the best arm (Russo, 2016; Qin et al., 2017; Shang et al., 2020).

☞ Asymptotic β-optimality: for all $\boldsymbol{F} \in \mathcal{F}^K$,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T^\star_\beta(\boldsymbol{F}) \,,$$

where $T^\star_\beta(\boldsymbol{F})$ same as $T^\star(\boldsymbol{F})$ with the constraint $w_{i^\star} = \beta$.

? 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 β-optimality

Sub-class of algorithms: β proportion of samples to the best arm (Russo, 2016; Qin et al., 2017; Shang et al., 2020).

☞ Asymptotic β-optimality: for all $\boldsymbol{F} \in \mathcal{F}^K$,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T^\star_\beta(\boldsymbol{F}) \,,$$

where $T^\star_\beta(\boldsymbol{F})$ same as $T^\star(\boldsymbol{F})$ with the constraint $w_{i^\star} = \beta$.

? 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 β-optimal instances.
3. Competitive performance on a real-world non-parametric task.

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 δ-correct algorithm ?

☞ calibrated **GLR stopping rule**

$$\tau_\delta = \inf\left\{ n \in \mathbb{N} \mid \min_{j\neq \hat{\imath}_n} W_n(\hat{\imath}_n, j) > \log\left(\frac{K-1}{\delta}\right) + 2\log\left(1 + n/2\right) + 2 \right\} \,,$$

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

$$W_n(i,j) = \inf_{x\in[0,B]} \left[N_{n,i} \mathcal{K}_{\inf}^-(F_{n,i}, x) + N_{n,j} \mathcal{K}_{\inf}^+(F_{n,j}, x) \right] \,.$$

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 δ-correct algorithm ?

☞ calibrated **GLR stopping rule**

$$\tau_\delta = \inf\left\{ n \in \mathbb{N} \mid \min_{j\neq\hat{\imath}_n} W_n(\hat{\imath}_n, j) > \log\left(\frac{K-1}{\delta}\right) + 2\log\left(1 + n/2\right) + 2 \right\} \,,$$

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

$$W_n(i,j) = \inf_{x\in[0,B]} \left[N_{n,i} \mathcal{K}_{\inf}^-(F_{n,i}, x) + N_{n,j} \mathcal{K}_{\inf}^+(F_{n,j}, x) \right] \,.$$

Sampling rule

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

Leader/Challenger

Choices of the 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 Π_{n-1} on $(0, B)^K$, $B_n^{\mathsf{TS}} \in \arg\max_{i \in [K]} \theta_i$ where $\theta \sim \Pi_{n-1}$.

Choices of the challenger given leader B_n:

- Transportation Cost [Improved] (**TC[I]**), deterministic, $C_n^{\mathsf{TC[I]}} \in \arg\min_{j \neq B_n} W_{n-1}(B_n, j)\ [+ \log N_{n-1,j}]$.
- Re-Sampling (**RS**), randomized, repeat $\theta \sim \Pi_{n-1}$ until $C_n^{\mathsf{RS}} \in \arg\max_{i \in [K]} \theta_i \not\ni B_n$.

Leader/Challenger

Choices of the 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 Π_{n-1} on $(0, B)^K$, $B_n^{\mathsf{TS}} \in \arg\max_{i \in [K]} \theta_i$ where $\theta \sim \Pi_{n-1}$.

Choices of the challenger given leader B_n:

- Transportation Cost [Improved] (**TC[I]**), deterministic, $C_n^{\mathsf{TC[I]}} \in \arg\min_{j \neq B_n} W_{n-1}(B_n, j)\ [+ \log N_{n-1,j}]$.
- Re-Sampling (**RS**), randomized, repeat $\theta \sim \Pi_{n-1}$ until $C_n^{\mathsf{RS}} \in \arg\max_{i \in [K]} \theta_i \not\ni B_n$.

Sample complexity upper bound

Theorem

Let $\beta \in (0,1)$. Instantiating the Top Two algorithm with any pair of leader/challenger introduced above yields a δ-correct algorithm and, for all $\boldsymbol{F} \in \mathcal{F}^K$ with $m(\boldsymbol{F}) \in (0,B)^K$ and $\min_{i \neq j} |m(F_i) - m(F_j)| > 0$,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T^\star_\beta(\boldsymbol{F}) \, .$$

Recommendations: β-EB-TCI and β-TS-TC.

Section 4

Experiments

Empirical results

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

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

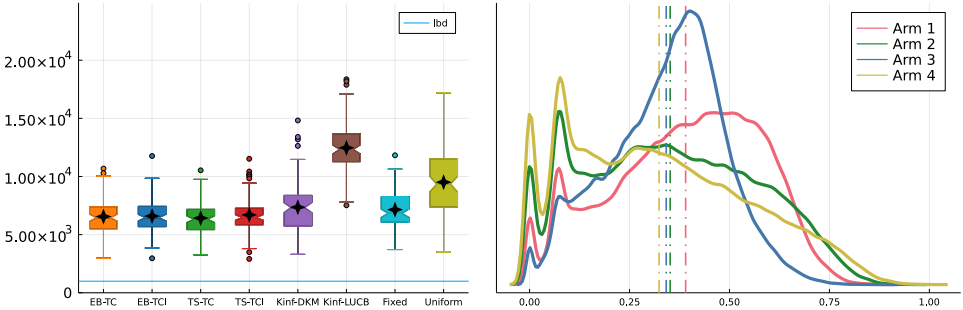

Figure: Empirical stopping time (a) on scaled DSSAT instances with their density and mean (b). Lower bound is $T^\star(\boldsymbol{F}) \log(1/\delta)$.

Conclusion

1. Generic and modular analysis of Top Two algorithms.
2. Asymptotically β-optimal instances.
3. Competitive performance on a real-world non-parametric task.

References

Agrawal, S., Juneja, S., and Glynn, P. W. (2020). Optimal δ-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)*.

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

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 ?

Computations

Computing transportation costs between arm i and arm j:

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

? "Posterior" sampling for bounded ? (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

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

Comparing instances

Limitations:

- For large n, the RS challenger is computationally costly and the TS leader is expensive.
- β-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: β-EB-TCI and β-TS-TC.

Random Bernoulli instances

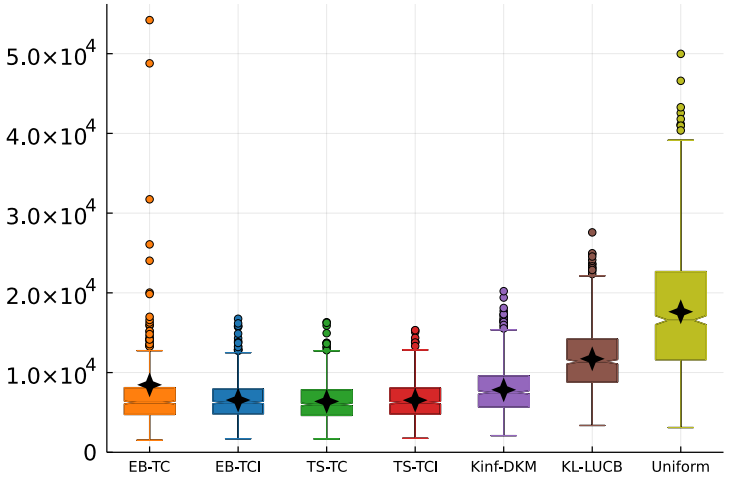

Figure: Empirical stopping time on random Bernoulli instances with $K = 10$.

Novelties

Literature:

- TTTS and T3C corresponds to β-TS-RS and β-TS-TC.
- β-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.

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.