Identifying the best treatment for a mixture of subpopulations

UT Seminar in honour of Stef Baas' PhD defence

Wouter M. Koolen

with Y. Russac, C. Katsimerou, D. Bohle, O. Cappé, A. Garivier

15 Nov 2024

Centrum Wiskunde & Informatica

UNIVERSITY OF TWENTE.

The Problem

Two treatments:

Control Developmental

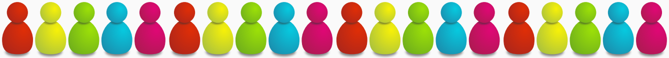
The Problem

Two treatments:

Control

Developmental

A stream of participants:



with sub-population identifier

What we want to know

Control Developmental

with sub-population identifier

Question of interest:

BAI Which of $\{C, D\}$ is the best overall treatment?

How does the presence of **sub-populations** affect learning?

Model for the Environment

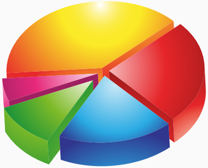
Definition (Natural Frequencies)

- $\boldsymbol{\alpha} \in \triangle_J$: frequency of the J subpopulations

Definition (Bandit)

A *bandit with* 2 *treatments and J subpopulations* is

- $\boldsymbol{\theta} \in [0,1]^{2 \times J}$: matrix of Bernoulli reward distributions



$\boldsymbol{\alpha}$

			...	
C	0.1	0.5	...	0.8
D	0.3	0.2	...	0.1

$\boldsymbol{\theta}$

Model for the Environment

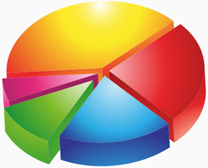
Definition (Natural Frequencies)

- $\alpha \in \triangle_J$: frequency of the J subpopulations

Definition (Bandit)

A *bandit with 2 treatments and J subpopulations* is

- $\theta \in [0,1]^{2 \times J}$: matrix of Bernoulli reward distributions



α

			...	
C	0.1	0.5	...	0.8
D	0.3	0.2	...	0.1

θ

Natural frequencies α are known and bandit θ is **unknown**.

The Target

Best Treatment Overall (BAI-S)

Given α, the correct answer for bandit θ is

$$i^*(\theta) = \underset{a \in \{C,D\}}{\operatorname{argmax}} \sum_{j=1}^{J} \alpha_j \theta_{a,j}$$

The Protocol

We study four **Modes of Interaction**

Protocol

for $t = 1, 2, \ldots$ **until** Learner decides to stop

-

Oblivious	Agnostic	Proport.	Active
Pick A_t	Pick A_t	See $J_t \sim \boldsymbol{\alpha}$	
Hidden $J_t \sim \boldsymbol{\alpha}$	See $J_t \sim \boldsymbol{\alpha}$	Pick A_t	Pick A_t and J_t

- See reward $X_t \sim \theta_{A_t, J_t}$

Learner recommends $\hat{\imath} \in \{C, D\}$ (best treatment)

Modes *constrain* the joint distribution of A_t and J_t

Goal and Approach

We seek a **response-adaptive** policy for Learner that

(1) is δ-PAC, i.e. for any bandit $\boldsymbol{\theta}$,

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\text{Learner stops and recommends wrong answer}\right) \leq \delta.$$

(2) minimises **sample complexity**, i.e. $\mathbb{E}_{\boldsymbol{\theta}}\left[\text{stopping time}\right]$

Our Results

(Russac, Katsimerou, Bohle, Cappé, Garivier, and Koolen, 2021)

- Information-theoretic lower bounds for all four modes
- Matching ($\delta \to 0$) algorithms (Track-and-Stop family)

Lower Bound

Theorem

For any policy, the expected number of rounds for the BAI-S problem on $\boldsymbol{\theta}$ with mode constraint $\mathcal{C}$ satisfies

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau_\delta]}{\ln(1/\delta)} \geq T^\star_{\mathcal{C}}(\boldsymbol{\theta})$$

where

$$T^\star_{\mathcal{C}}(\boldsymbol{\theta})^{-1} = \max_{\boldsymbol{w} \in \mathcal{C}} \inf_{\substack{\boldsymbol{\lambda} \in [0,1]^{2 \times J} \\ \boldsymbol{\alpha}^\intercal \boldsymbol{\lambda}_C = \boldsymbol{\alpha}^\intercal \boldsymbol{\lambda}_D}} \sum_{a \in \{C,D\}} \sum_{i=1}^{J} w_{a,i} \, \mathrm{KL}(\theta_{a,i}, \lambda_{a,i})$$

NB: the $\min/\inf$ is the (expected) amount of statistical evidence collected per round by sampling proportions $\boldsymbol{w}$ against any bandit $\boldsymbol{\lambda}$ with $i^*(\boldsymbol{\lambda}) \neq i^*(\boldsymbol{\theta})$

Upper Bound Intuition

Estimate for treatment quality carries **uncertainty**:

$$\sum_{j=1}^{J} \alpha_j \hat{\theta}_{a,j}$$

Uncertainty $\Leftrightarrow$ variance.

If each arm a, j of variance $\sigma^2_{a,j} = \theta_{a,j}(1 - \theta_{a,j})$ is sampled $n_{a,j}$ times

$$\mathbb{V}\left[\sum_{j=1}^{J} \alpha_j \hat{\theta}_{a,j}\right] = \sum_{j=1}^{J} \alpha_j^2 \mathbb{V}\left[\hat{\theta}_{a,j}\right] = \sum_{j=1}^{J} \frac{\alpha_j^2 \sigma^2_{a,j}}{n_{a,j}}$$

Upper Bound Intuition

Estimate for treatment quality carries **uncertainty**:

$$\sum_{j=1}^{J} \alpha_j \hat{\theta}_{a,j}$$

Uncertainty $\Leftrightarrow$ variance.

If each arm a, j of variance $\sigma_{a,j}^2 = \theta_{a,j}(1 - \theta_{a,j})$ is sampled $n_{a,j}$ times

$$\mathbb{V}\left[\sum_{j=1}^{J} \alpha_j \hat{\theta}_{a,j}\right] = \sum_{j=1}^{J} \alpha_j^2 \mathbb{V}\left[\hat{\theta}_{a,j}\right] = \sum_{j=1}^{J} \frac{\alpha_j^2 \sigma_{a,j}^2}{n_{a,j}}$$

Minimised unconstrained (active mode) at

$$n_{a,j} \propto \alpha_j \sigma_{a,j}$$

Other modes: add **constraints** $\boldsymbol{n} \in \mathcal{C}$

Results (explicit Gaussian case)

Denoting the gap by $\Delta = \sum_{j=1}^{J} \alpha_j(\theta_{C,j} - \theta_{D,j})$, we find

better ↓

$$T^\star_{\text{oblivious}}(\boldsymbol{\theta}) \approx \frac{2\left(\sum_{a\in\{C,D\}} \sqrt{\sum_{j=1}^{J} \alpha_j\left(\sigma_{a,j}^2 + (\theta_{a,j} - \boldsymbol{\alpha}^\intercal \boldsymbol{\theta}_a)^2\right)}\right)^2}{\Delta^2}$$

$$T^\star_{\text{agnostic}}(\boldsymbol{\theta}) = \frac{2\left(\sqrt{\sum_{j=1}^{J} \alpha_j \sigma_{C,j}^2} + \sqrt{\sum_{j=1}^{J} \alpha_j \sigma_{D,j}^2}\right)^2}{\Delta^2}$$

$$T^\star_{\text{proport.}}(\boldsymbol{\theta}) = \frac{2\sum_{j=1}^{J} \alpha_j(\sigma_{C,j} + \sigma_{D,j})^2}{\Delta^2},$$

$$T^\star_{\text{active}}(\boldsymbol{\theta}) = \frac{2\left(\sum_{j=1}^{J} \alpha_j(\sigma_{C,j} + \sigma_{D,j})\right)^2}{\Delta^2},$$

Algorithm

Sampling Rule

Ensure that actual sampling proportions $\boldsymbol{N}_t/t$ track oracle proportions at plug-in estimate $\hat{\theta}(t)$

$$\boldsymbol{w}^*(\hat{\theta}(t)) = \arg\max_{\boldsymbol{w}\in\mathcal{C}} \inf_{\substack{\boldsymbol{\lambda}\in[0,1]^{2\times J} \\ \boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_C=\boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_D}} \sum_{a\in\{C,D\}} \sum_{j=1}^{J} w_{a,j}\,\mathrm{KL}(\hat{\theta}_{a,j}(t), \lambda_{a,j})$$

Tracking is done locally, respecting the mode *constraint*

Stopping Rule (GLRT)

Stop at $\tau_\delta = t$ when we've collected enough information, i.e.

$$\inf_{\substack{\boldsymbol{\lambda}\in[0,1]^{2\times J} \\ \boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_C=\boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_D}} \sum_{a\in\{C,D\}} \sum_{j=1}^{J} N_{a,j}(t)\,\mathrm{KL}(\hat{\theta}_{a,j}(t), \lambda_{a,j}) \geq \ln\frac{\ln t}{\delta}$$

Recommendation Rule

Output $i^*(\hat{\theta}(t))$

Validation: Asymptotic Optimality

Theorem

The stopping+recommendation rules are δ-PAC.

Theorem

The algorithm ensures that the expected number of rounds for the BAI-S problem with mode constraint $\mathcal{C}$ satisfies

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau_\delta]}{\ln(1/\delta)} \leq T^\star_{\mathcal{C}}(\boldsymbol{\theta})$$

Upper bound matching lower bound, perfectly.

Conclusion

- Subpopulation awareness reduces **sample complexity** . . .
  . . . even if only interested in overall best treatment!

Conclusion

- Subpopulation awareness reduces **sample complexity** ...
  ...even if only interested in overall best treatment!

Start of a journey:

- Going beyond asymptotic optimality
- Structured (shape-constrained) mean matrices
- (Non)-parametric reward models
- $\epsilon > 0$

Thanks!

References

Garivier, A. and E. Kaufmann (2016). **"Optimal best arm identification with fixed confidence".** In: *Conference on Learning Theory*. PMLR, pp. 998–1027.

Russac, Y., C. Katsimerou, D. Bohle, O. Cappé, A. Garivier, and W. M. Koolen (Dec. 2021). **"A/B/n Testing with Control in the Presence of Subpopulations".** In: *Advances in Neural Information Processing Systems (NeurIPS) 34*. Accepted.