# A/B/n Testing with Control in the Presence of Subpopulations

SCOOL Seminar, INRIA Lille

**Wouter M. Koolen**
with Y. Russac, C. Katsimerou, D. Bohle, O. Cappé, A. Garivier
28 Jan 2022

CWI
Centrum Wiskunde & Informatica

# Meta-level questions

## Pure Exploration

- We want algorithms for adaptively choosing **experiments**
- in an a-priori **unknown environment**
- to gain information as fast as possible
- about which decision is **correct/best/useful**

# Meta-level questions

## Pure Exploration

- We want algorithms for adaptively choosing **experiments**
- in an a-priori **unknown environment**
- to gain information as fast as possible
- about which decision is **correct/best/useful**

$\Rightarrow$ Active Sequential Composite Multiple Hypothesis Test

# Meta-level questions

**Pure Exploration**

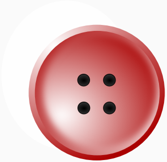
- We want algorithms for adaptively choosing **experiments**
- in an a-priori **unknown environment**
- to gain information as fast as possible
- about which decision is **correct/best/useful**

$\Rightarrow$ Active Sequential Composite Multiple Hypothesis Test

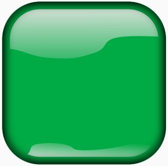
**Examples**

- $(\epsilon, \delta)$-PAC learning of policy in unknown MDP
- (MCTS) simulation-based planning
- Drug safety/efficacy trial
- This talk: A/B/n testing with control

# The Problem



Control



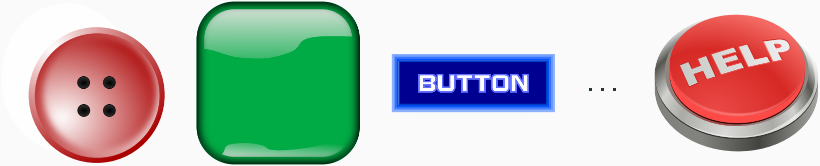
Version 1


Version 2

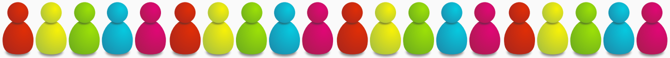
. . .


Version $K$

# The Problem




Control Version 1 Version 2 ... Version $K$



# The Question


Many questions interesting:

| | | | |
|---|---|---|---|
| | BAI | What is the best version? | $\{0, \dots, K\}$ |
| | Thr. | Which versions are at least $\theta$-good? | $\mathcal{P}(\{0, \dots, K\})$ |
| | • | Is any version better than the control? | $\{\text{yes}, \text{no}\}$ |
| | • | Which version, if any, beats the control? | $\{0, \dots, K\}$ |
| this talk | ABC | Which versions are better than the control? | $\mathcal{P}(\{1, \dots, K\})$ |

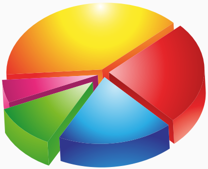
How does the presence of **sub-populations** affect learning?

# The Model

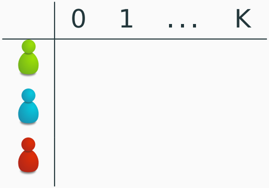
**Definition (Model)**

A *bandit with $K+1$ arms and $J$ subpopulations* is

- Distribution $\boldsymbol{\alpha}$ of the $J$ subpopulations
- a $(K+1) \times J$ matrix $\boldsymbol{\mu}$ of (Bernoulli, say) reward distributions



$\boldsymbol{\alpha}$




$\boldsymbol{\mu}$

# The Target

**All arms Better than the Control with Subpopulations (ABC-S)**

The correct answer for Bandit $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ is

$$S^*(\boldsymbol{\mu}) \;=\; \left\{k \in \{1, \ldots, K\} \,\middle|\, \sum_{j=1}^{J} \alpha_j \mu_{k,j} > \sum_{j=1}^{J} \alpha_j \mu_{0,j}\right\}$$

# 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 $I_t \sim \alpha$ | |
| Hidden $I_t \sim \alpha$ | See $I_t \sim \alpha$ | Pick $A_t$ | Pick $A_t$ and $I_t$ |

- See reward $X_t \sim \mu_{A_t, I_t}$

Learner recommends $\hat{S} \subseteq \{1, \ldots, K\}$ (arms better than control)

Modes *constrain* the sampling proportions of $(I, A)$

# Goal and Approach

We want our learner to

(1) be $\delta$-PAC, i.e. for any bandit $\boldsymbol{\mu}$,

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

(2) minimise **sample complexity**, i.e. $\mathbb{E}_{\boldsymbol{\mu}}[\text{stopping time}]$

# 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)

# Results

Let's think about $K = 1$ arm vs control with $J$ subpopulations.

Let's investigate the **Gaussian** case: reward for $a, j$ is $\mathcal{N}(\mu_{a,j}, \sigma^2_{a,j})$.

# Lower Bound

**Theorem**

*For any strategy, the expected number of rounds for the ABC-S problem with mode constraint $\mathcal{C}$ satisfies*

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

*where*

$$\begin{aligned} T^\star(\boldsymbol{\mu})^{-1} &= \max_{\boldsymbol{w} \in \mathcal{C}} \inf_{\boldsymbol{\lambda} : S^*(\boldsymbol{\lambda}) \neq S^*(\boldsymbol{\mu})} \sum_{a=0}^{K} \sum_{i=1}^{J} w_{a,i} \, \mathrm{KL}(\mu_{a,i}, \lambda_{a,i}) \\ &= \max_{\boldsymbol{w} \in \mathcal{C}} \min_{b \neq 0} \inf_{\substack{\boldsymbol{\lambda} \in \mathcal{L} \\ \boldsymbol{\alpha}^\intercal \boldsymbol{\lambda}_0 = \boldsymbol{\alpha}^\intercal \boldsymbol{\lambda}_b}} \sum_{a \in \{0,b\}} \sum_{i=1}^{J} w_{a,i} \, \mathrm{KL}(\mu_{a,i}, \lambda_{a,i}) \end{aligned}$$

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 $S^*(\boldsymbol{\lambda}) \neq S^*(\boldsymbol{\mu})$

# How to think about this

Estimate for arm quality carries **uncertainty**:

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

Uncertainty $\Leftrightarrow$ variance. Now if $(a, j)$ is sampled $n_{a,j}$ times,

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

# How to think about this

Estimate for arm quality carries **uncertainty**:

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

Uncertainty $\Leftrightarrow$ variance. Now if $(a, j)$ is sampled $n_{a,j}$ times,

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

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)

$K = 1$ arm vs control with $J$ subpopulations.

Denoting the gap by $\Delta_1 = \sum_{j=1}^{J} \alpha_j(\mu_{1,j} - \mu_{0,j})$, we find

better

$$
\begin{aligned}
T^\star_{\text{oblivious}}(\boldsymbol{\mu}) \;&\approx\; \frac{2\left(\sum_{a\in\{0,1\}} \sqrt{\sum_{j=1}^{J} \alpha_j(\sigma_{a,j}^2 + (\mu_{a,j} - \mu_a)^2)}\right)^2}{\Delta_1^2} \\
T^\star_{\text{agnostic}}(\boldsymbol{\mu}) \;&=\; \frac{2\left(\sqrt{\sum_{j=1}^{J} \alpha_j \sigma_{0,j}^2} + \sqrt{\sum_{j=1}^{J} \alpha_j \sigma_{1,j}^2}\right)^2}{\Delta_1^2} \\
T^\star_{\text{proport.}}(\boldsymbol{\mu}) \;&=\; \frac{2\sum_{j=1}^{J} \alpha_j(\sigma_{0,j} + \sigma_{1,j})^2}{\Delta_1^2}, \\
T^\star_{\text{active}}(\boldsymbol{\mu}) \;&=\; \frac{2\left(\sum_{j=1}^{J} \alpha_j(\sigma_{0,j} + \sigma_{1,j})\right)^2}{\Delta_1^2},
\end{aligned}
$$

# Algorithm

### Sampling Rule

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

$$\boldsymbol{w}^*(\hat{\boldsymbol{\mu}}(t)) \;=\; \arg\max_{\boldsymbol{w}\in\mathcal{C}} \min_{b\neq 0} \inf_{\substack{\boldsymbol{\lambda}\in\mathcal{L} \\ \boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_0=\boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_b}} \sum_{a\in\{0,b\}} \sum_{i=1}^{J} w_{a,i}\,\mathrm{KL}(\hat{\mu}_{a,i}(t),\lambda_{a,i})$$

Tracking is done locally, respecting the mode *constraint*

### Stopping Rule

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

$$\min_{b\neq 0} \inf_{\substack{\boldsymbol{\lambda}\in\mathcal{L} \\ \boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_0=\boldsymbol{\alpha}^\intercal\boldsymbol{\lambda}_b}} \sum_{a\in\{0,b\}} \sum_{i=1}^{J} N_{a,i}(t)\,\mathrm{KL}(\hat{\mu}_{a,i}(t),\lambda_{a,i}) \;\geq\; \ln\frac{\ln t}{\delta}$$

### Recommendation Rule

Output $S^*(\hat{\boldsymbol{\mu}}(t))$

# Validation: Asymptotic Optimality

**Theorem**

*The stopping+recommendation rules are $\delta$-PAC.*

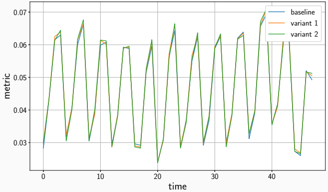
**Theorem**

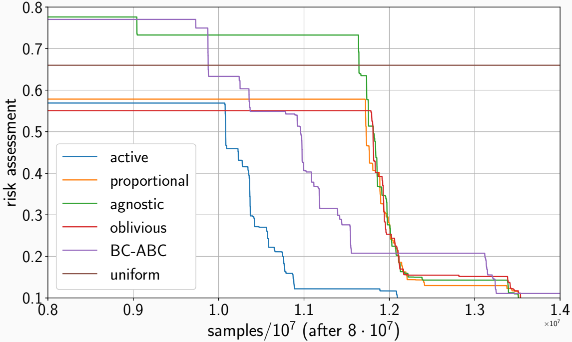
*The algorithm ensures that the expected number of rounds for the ABC-S problem with mode constraint $\mathcal{C}$ satisfies*

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

Upper bound matching lower bound, perfectly.

# Validation: Real-world experiment with $\delta = 0.1$



# Conclusion

- Interesting pure exploration problems in A/B/n testing
- Subpopulation awareness reduces **sample complexity** . . .
  . . . even if the recommendations (arms) are unaware!

## Conclusion

- Interesting pure exploration problems in A/B/n testing
- Subpopulation awareness reduces **sample complexity** ...
  ...even if the recommendations (arms) are unaware!

Next steps:

- Going beyond asymptotic optimality
- Structured (shape-constrained) mean matrices

# 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.