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

CWI Group Meeting

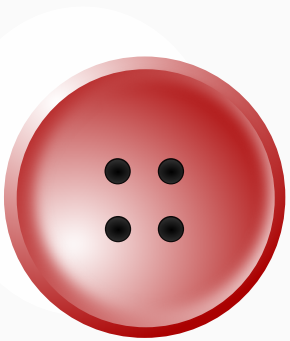
Wouter M. Koolen

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

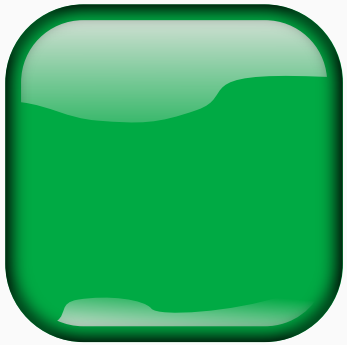
15 Oct 2021

Centrum Wiskunde & Informatica

The Problem



Control



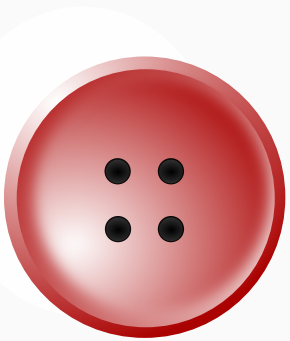
Version 1

Version 2

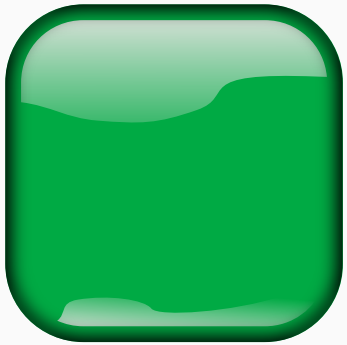
. . .

Version *K*

The Problem



Control



Version 1

Version 2

. . .

Version *K*

The Question

Many questions interesting:

- What is the best version? $\{0, \ldots, K\}$
- Is any version better than the control? $\{\text{yes}, \text{no}\}$
- Which version, if any, is better than the control? $\{0, \ldots, K\}$
- Which versions are better than the control? $\mathcal{P}(K)$

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

The Protocol

Definition (Model)

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

- a $(K+1) \times J$ matrix μ of (Bernoulli, say) reward distributions
- distribution α of the J subpopulations

The Protocol

Definition (Model)

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

- a $(K+1) \times J$ matrix μ of (Bernoulli, say) reward distributions
- distribution α of the J subpopulations

The correct answer for (μ, α) is

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

The Protocol

Definition (Model)

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

- a $(K+1) \times J$ matrix μ of (Bernoulli, say) reward distributions
- distribution α of the J subpopulations

The correct answer for (μ, α) is

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

Protocol for four Modes of Interaction

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)

Goal and Approach

We want our learner to

(1) be δ-PAC, i.e. for any bandit μ,

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

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

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)
- Running risk assessment

Results

Let's think about one-arm vs control with J subpopulations.

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

How to think about this

Estimates for arm qualities are **uncertain**:

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

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

Estimates for arm qualities are **uncertain**:

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

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

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 ↓

$$T^\star_{\text{oblivious}}(\boldsymbol{\mu}) \approx \frac{2\left(\sum_{a\in\{0,1\}} \sqrt{\sum_{j=1}^{J} \alpha_j(\sigma^2_{a,j} + (\mu_{a,j} - \mu_a)^2)}\right)^2}{\Delta_1^2}$$

$$T^\star_{\text{agnostic}}(\boldsymbol{\mu}) = \frac{2\left(\sqrt{\sum_{j=1}^{J} \alpha_j \sigma^2_{0,j}} + \sqrt{\sum_{j=1}^{J} \alpha_j \sigma^2_{1,j}}\right)^2}{\Delta_1^2}$$

$$T^\star_{\text{proport.}}(\boldsymbol{\mu}) = \frac{2\sum_{j=1}^{J} \alpha_j(\sigma_{0,j} + \sigma_{1,j})^2}{\Delta_1^2},$$

$$T^\star_{\text{active}}(\boldsymbol{\mu}) = \frac{2\left(\sum_{j=1}^{J} \alpha_j(\sigma_{0,j} + \sigma_{1,j})\right)^2}{\Delta_1^2},$$

Theory

Theorem

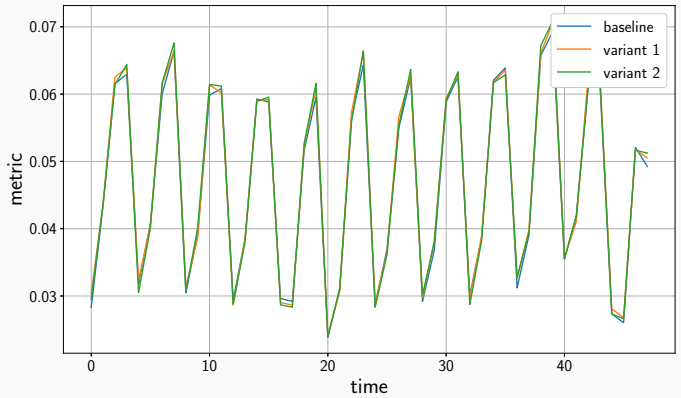
For any strategy, the expected number of rounds for the ABC-S problem satisfies

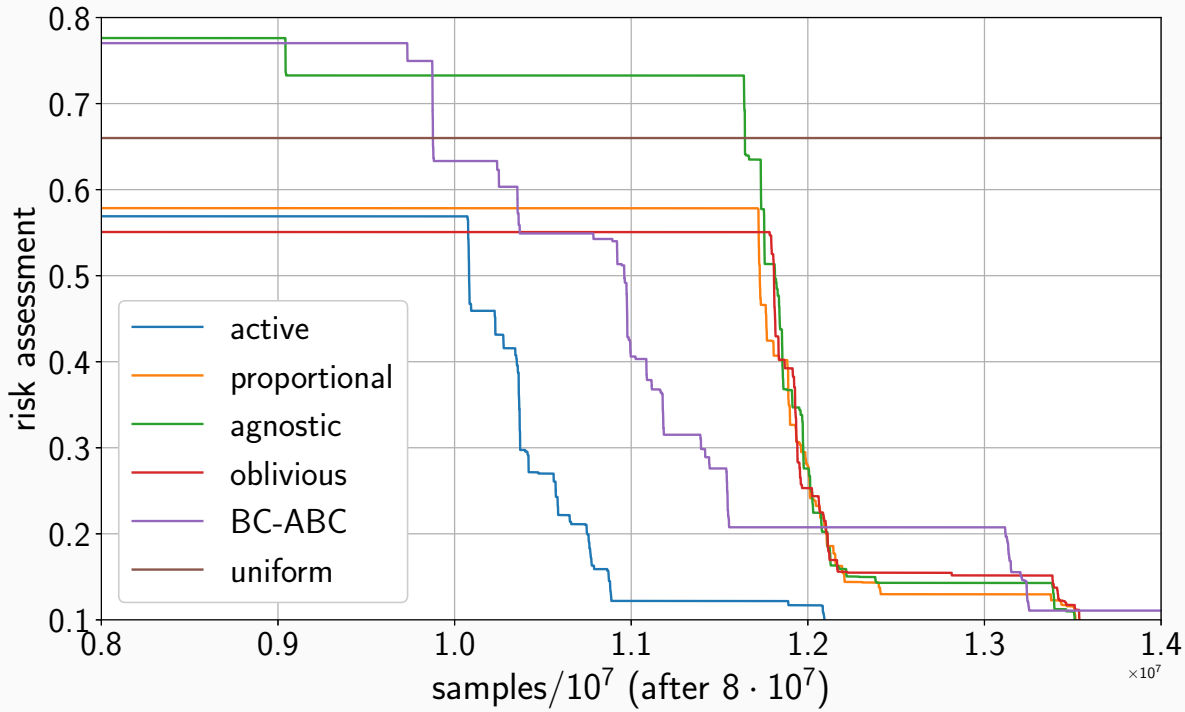
$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} \geq T^\star(\boldsymbol{\mu}) \tag{1}$$

where

$$T^\star(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{w} \in \mathcal{C}} \min_{b \neq 0} \inf_{\boldsymbol{\lambda} \in \mathcal{L}: \lambda_0 < \lambda_b} \sum_{a \in \{0,b\}} \sum_{i=1}^{J} w_{a,i} d(\mu_{a,i}, \lambda_{a,i}) \tag{2}$$

Real-world experiment





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.