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

Yoan Russac[1], Christina Katsimerou[2], Dennis Bohle[2],
Olivier Cappé[1], Aurélien Garivier[3] & Wouter M. Koolen[4]
[1] CNRS & ENS, [2] Booking.com, [3] ENS Lyon, [4] CWI

Problem

- A/B/n testing compares multiple website versions (called *arms*) to determine the one with the highest conversion.
- Online firms deploy the arms that satisfy multiple constraints (cost, strategy, etc.), as long as it is better than the baseline, the *control arm*.
- All Arms Better than the Control (**ABC**)
  $\neq$ Best arm identification (**BAI**)
  $\neq$ Better than a Threshold

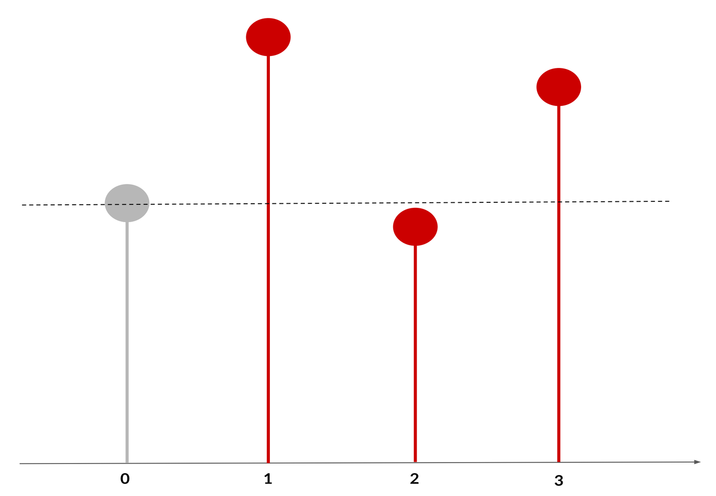


Figure 1: For the ABC problem we need to sample more arms 0 and 2, for the BAI problem we need to sample more 1 and 3 and for thresholding bandit we need to sample more 2 (control mean is known).

Challenge

Limitations of conventional A/B/n:
1. Uniform allocation of options to users is inefficient
2. Pre-determined experiment duration can be conservative

We aim to optimise adaptively:
1. the allocation of options to users
2. the stopping time of the A/B/n experiment

→ Traditional stochastic bandits assume that the arm samples are i.i.d., whereas real world data exhibit inhomogeneity, for instance seasonality patterns.

Objective

Identify the set of *Arms* that are *Better* than the *Control* in the presence of *Sub-populations* (ABC-S):

$$\mathcal{S}_\beta(\boldsymbol{\mu}) = \left\{ a \in \{1,\dots,K\} \text{ s.t } \sum_{i=1}^J \beta_i \mu_{a,i} > \sum_{i=1}^J \beta_i \mu_{0,i} \right\} ,$$

in the *fixed confidence* setting, i.e. for any *risk level* δ the probability of returning an incorrect answer must be $\leq \delta$.

The user at time t belongs to a *subpopulation* $I_t \in \{1,\dots,J\}$

- α_i is the natural proportion of subpopulation i
- $\mu_{a,i}$ is the mean reward of arm a for the i-th subpopulation
- $\boldsymbol{\beta} = (\beta_i)_{i=1,\dots,J}$ are *known user-defined population weights* defining the value of an arm

$$\mu_a = \sum_{i=1}^J \beta_i \mu_{a,i} .$$

Different *modes of interaction* with the subpopulations

Oblivious:
1. Pick A_t
2. Don't see $I_t \sim \boldsymbol{\alpha}$
3. See $X_t \sim \nu_{A_t,I_t}$

Agnostic:
1. Pick A_t
2. See $I_t \sim \boldsymbol{\alpha}$
3. See $X_t \sim \nu_{A_t,I_t}$

Proportional:
1. See $I_t \sim \boldsymbol{\alpha}$
2. Pick A_t
3. See $X_t \sim \nu_{A_t,I_t}$

Active:
1. Pick A_t and I_t
2. See $X_t \sim \nu_{A_t,I_t}$

Theoretical guarantees

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

Complexity of the learning problems

By remarking that $\mathcal{C}_{\text{agnostic}} \subset \mathcal{C}_{\text{prop}} \subset \mathcal{C}_{\text{active}}$, it holds that

$$\forall \boldsymbol{\mu} \in \mathcal{L}, \quad T^\star_{\text{active}}(\boldsymbol{\mu}) \leq T^\star_{\text{proportional}}(\boldsymbol{\mu}) \leq T^\star_{\text{agnostic}}(\boldsymbol{\mu}) . \tag{2}$$

When $\boldsymbol{\alpha} = \boldsymbol{\beta}$, for a *safely calibrated* oblivious policy, we further have

$$\forall \boldsymbol{\mu} \in \mathcal{L}, \quad T^\star_{\text{agnostic}}(\boldsymbol{\mu}) \leq T^\star_{\text{oblivious}}(\boldsymbol{\mu}) . \tag{3}$$

Track-and-Stop Algorithm

For $t \geq 1$:

- Sampling rule: given the current estimates
  1. estimate the target weights $\mathbf{w}_t$
  2. pick arm $\begin{cases} \text{active:} & (A_t, I_t) \in \text{argmax}_{a,i} N_{a,i}(t-1) - t\mathbf{w}_t(a,i) \\ \text{proportional:} & A_t \in \text{argmax}_a N_{a,I_t}(t-1) - t\alpha_{I_t}\mathbf{w}_t(a|I_t) \\ \text{agnostic:} & A_t \in \text{argmax}_a N_a(t-1) - t\mathbf{w}_t(a) \end{cases}$

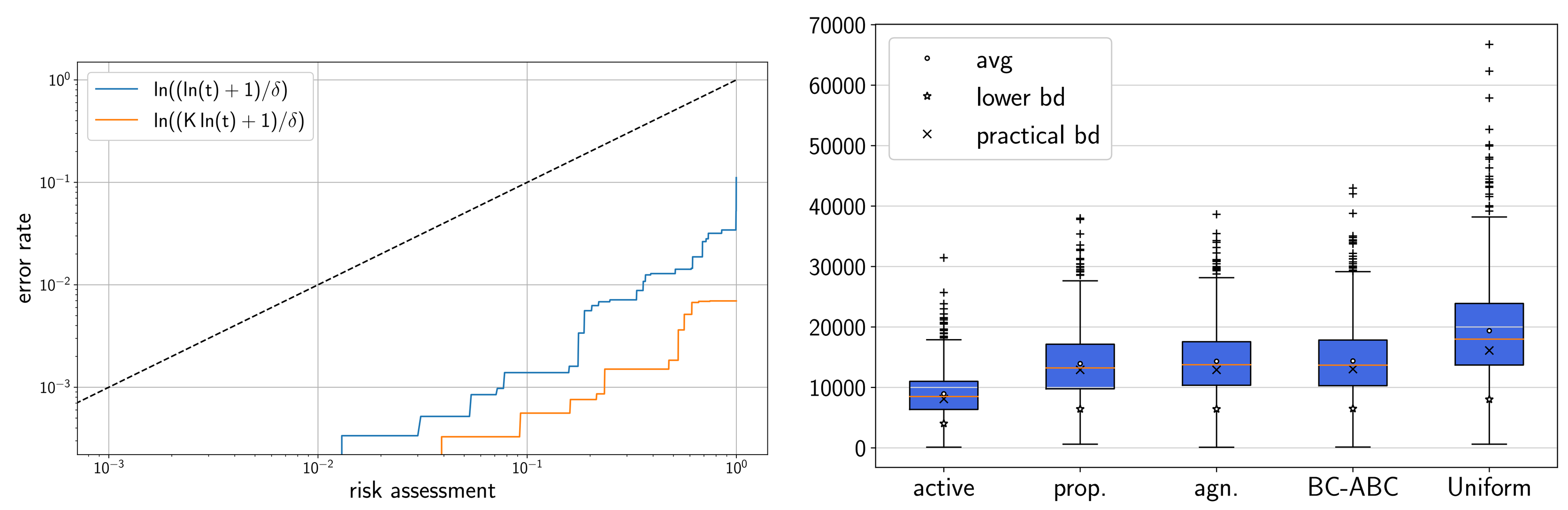
- Recommendation: $\mathcal{S}(\hat{\boldsymbol{\mu}}_t) = \{a \in \{1,\dots,K\} : \hat{\mu}_a(t) > \hat{\mu}_0(t)\}$ at confidence level $\hat{\delta}_t = \min\{\delta \in (0,1) | \Lambda(t) \geq \beta(t,\delta)\}$, obtained by inverting the threshold $\beta(t,\delta)$ at the GLR statistic

$$\Lambda(t) = \min_{b\neq 0} \inf_{\lambda\in\mathcal{L}:\lambda_0=\lambda_b} \sum_{a\in\{0,b\}} \sum_{i=1}^J N_{a,i}(t) d(\hat{\mu}_{a,i}(t), \lambda_{a,i}) . \tag{4}$$

- Calibration: For $\beta(t,\delta) = 6J \ln\ln t + \ln\frac{1}{\delta} + K + 2J \cdot O(\ln\ln\frac{1}{\delta})$, Track-and-Stop is *safely calibrated*:

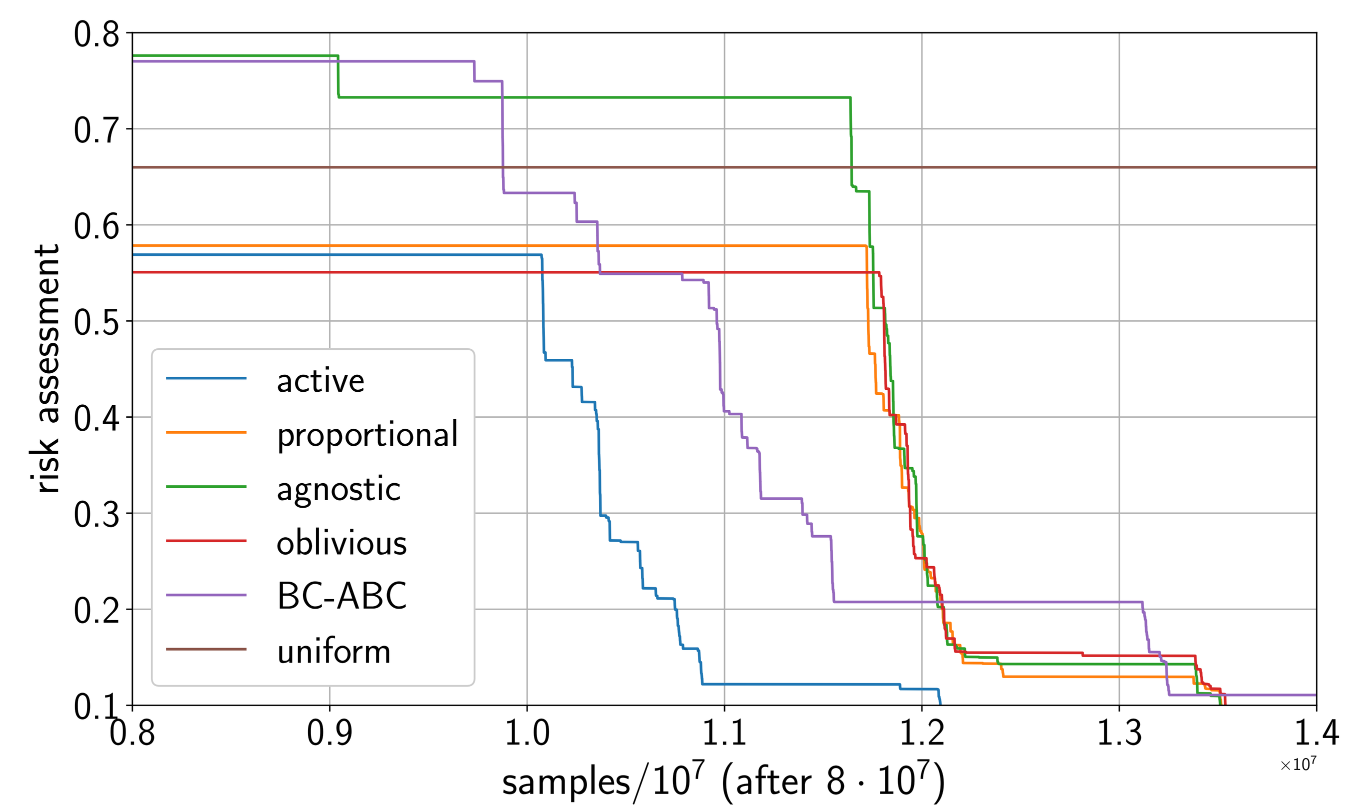
$$\forall \boldsymbol{\mu} \in \mathcal{L},\ \forall \delta \in (0,1), \quad \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t \geq 1 : \hat{\mathcal{S}}_t \neq \mathcal{S}(\boldsymbol{\mu}) \cap \hat{\delta}_t \leq \delta\right) \leq \delta . \tag{5}$$

Numerical Results



(Left) Risk assessment calibration on a log-log scale. In practice the threshold $\ln((1+\ln t)/\delta)$ works well. (Right) Stopping time boxplot for $\boldsymbol{\mu} = [0.1\ 0.4\ 0.3; 0.2\ 0.5\ 0.2; 0.5\ 0.1\ 0.1]$ when $\boldsymbol{\beta} = [1/3, 1/3, 1/3]$, $\boldsymbol{\alpha} = [0.4, 0.5, 0.1]$ with Bernoulli distributions.

Real Data Experiment: Booking.com webpage data



The experiment compares $K = 2$ copies of a component of the webpage against the baseline. Both copies are better than the control. Due to global traffic, the data exhibits seasonality patterns within a day. We treat the $J = 4$ seasons as i.i.d. subpopulations.