Batched fixed-confidence pure exploration for bandits with switching constraints

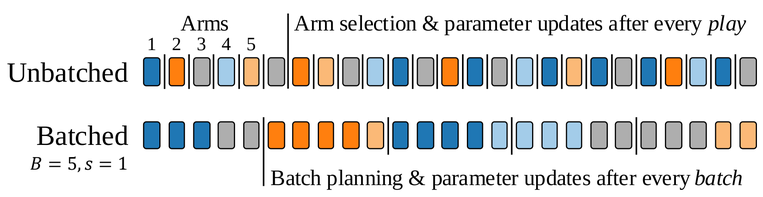
Newton Mwai, Milad Malekipirbazari, Fredrik D. Johansson
Computer Science and Engineering Department
Chalmers University of Technology, Sweden

CHALMERS UNIVERSITY OF TECHNOLOGY

Motivation & Research Goals

- Switching costs arise in real-world applications such as personalized medicine, in which changes in treatment may require a wash-out period where the patient is not taking any drug; or in industrial applications where reconfiguring production is costly.
- Controlling for switching is significantly understudied outside of regret minimization.
- In this work, we present a MAB formulation with constraints on the arm switching frequency in fixed-confidence pure exploration, by batching plays, give lower bound for this setting and present tracking algorithms with a limited number of arm switches.

Problem Formulation



Our goal is to design a search strategy ϕ ***to minimize the expected number of trials*** τ ***required to identify an optimal arm with confidence at least*** $1-\delta$ ***for a given*** $\delta > 0$***, while limiting the expected rate of switching arms to*** $\alpha \in [0,1]$ (Objective (1)).

Objective: Limiting Switching in PE

$$\begin{aligned} \underset{\phi}{\text{minimize}} \quad & \mathbb{E}_\phi[\tau] \\ \text{subject to} \quad & \mathbb{P}(\mu_{\hat{a}_\tau} < \mu^*) \leq \delta \\ & \mathbb{E}_\phi[S_\tau] \leq \alpha \mathbb{E}_\phi[\tau] \end{aligned} \tag{1}$$

Objective: Limiting switches in batches

$$\begin{aligned} \underset{\phi}{\text{minimize}} \quad & \mathbb{E}_\phi[\beta] \\ \text{subject to} \quad & \mathbb{P}\left(\mu_{\hat{a}_\beta} < \mu^*\right) \leq \delta \\ & S^b \leq s, \ \forall b \in \mathbb{N} \end{aligned} \tag{2}$$

We re-formulate our goal to be ***to minimize the expected number of batches*** β ***required to identify an optimal arm, with confidence at least*** $1-\delta$***, while limiting the arm switches within the batch to be at most a pre-specified switching constraint*** $s \in \{0, ..., \min(K-1, B-1)\}$ (Objective (2)).

Sparse batch proportions*:** We can determine sparsity-constrained integer playing ***batch configurations $\mathcal{C}^K_{B,s}$ of arm plays in the batch, ensuring that the plays matches the desired sparsity. Typically very high-dimensional, scaling exponentially with the number of arms.

Lower Bound: Let $\Sigma^{\mathcal{C}} := \Sigma^{|\mathcal{C}^K_{B,s}|-1}$ be the simplex over batch configurations of size B that use fewer than s switches. Given a confidence level $\delta \in (0,1)$, for any algorithm that returns the best arm with probability at least $1-\delta$, and for any bandit problem $\mu \in \mathbb{R}^K$, the following inequality holds:

$$\mathbb{E}_\mu[\beta] \geq T^*_{bc}(\mu) \cdot \mathrm{kl}(\delta, 1-\delta), \tag{3}$$

where the characteristic time $T^*_{bc}(\mu)$ is given by

$$T^*_{bc}(\mu)^{-1} := \sup_{p \in \Sigma^{\mathcal{C}}} \inf_{\lambda \in \mathrm{Alt}(\mu)} \sum_{a=1}^{K} \sum_{c \in \mathcal{C}^K_{B,s}} p_c c_a d(\mu_a, \lambda_a). \tag{4}$$

Tracking algorithms Garivier and Kaufmann (2016) introduced the idea of *track-and-stop* algorithms, designed to *track* the optimal arm playing proportions $w^*(\hat{\mu})$ of the lower bound,

$$w^*(\hat{\mu}) := \arg\max_{w \in \Sigma^K} \inf_{\lambda \in \mathrm{Alt}(\hat{\mu})} \left(\sum_{a=1}^{K} w_a d(\hat{\mu}_a, \lambda_a) \right). \tag{5}$$

Observation: If there exist configuration proportions $p^* \in \Sigma^{\mathcal{C}}$ such that $\sum_{c \in \mathcal{C}} p^*_c c_a = w^*(\hat{\mu})$ for w^* in Eq. (5), these are minimizers of Eq. (4)

When tracking, we aim to minimize the total positive deficit between expected plays and actual plays: $(d_a(b))_+ := (B\bar{w}_a(b) - N_a(b))_+$

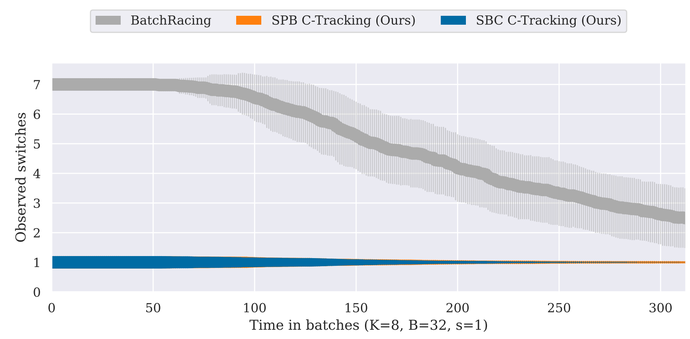
Selected Results

Algorithm 1: Sparse Batch Configurations (SBC) and Sparse Projected Batch (SPB) C-Tracking

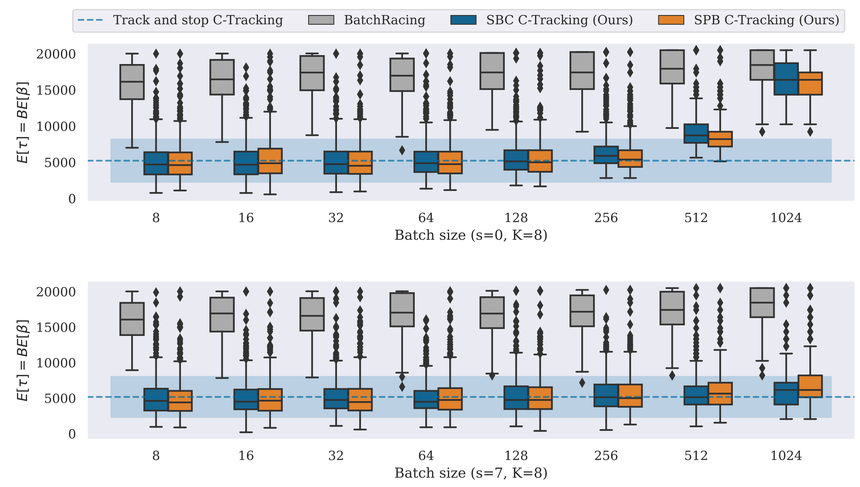
Input: K arms, $\delta \in (0,1)$, B: batch size, s: batch switch limit
Output: β, $\hat{a}_\beta$

$b \leftarrow 1$, $t \leftarrow 1$, $Z_1 \leftarrow 0$, $\hat{\mu}_0 \leftarrow \mathbf{0}$, $N(1) \leftarrow \mathbf{0} \in \mathbb{R}^K$;
while $Z_b \leq \log\left(\frac{\log(bB)+1}{\delta}\right)$ **do**
  Let $\epsilon_b \leftarrow (K^2 + bB)^{-1/2}/2$;
  Compute $w^{\epsilon_b - 1}(\hat{\mu}_{b-1})$;
  Compute $d(b) = B\sum_{i=0}^{b-1} w^{\epsilon_i}(\hat{\mu}_i) - N(b)$;
  if *SBC C-Tracking* **then**
    Let $\tilde{c} \in \arg\min_{c \in \mathcal{C}^K_{B,s}} \sum_{a=1}^{K} (d_a(b) - c_a)_+$;
    Greedy batch filling;
  else
    Let $\hat{w}^{s+1}(b) \in \arg\min_{w \in \sum^K_{s+1}} \|w - (\bar{d}(b))_+\|_2$;
    and $\tilde{c} = \text{integer}(\hat{w}^{s+1} * B)$;
    Proportional filling (SPB C-Tracking);
  while $t \leq bB$ **do**
    Let $\bar{a} \leftarrow \arg\max_{a \in \mathcal{A}} \tilde{c}$ and $\bar{c} = \tilde{c}_{\bar{a}}$;
    Play $a_t, ..., a_{t+\bar{c}-1}$ with arm $\bar{a}$;
    Observe rewards $(r_t, ..., r_{t+\bar{c}-1})$;
    $N_{a_t}(b+1) \leftarrow N_{a_t}(b) + \bar{c}$;
    $\hat{\mu}_{b,a_t} \leftarrow \frac{1}{N_{\bar{a}}(b+1)} \sum_{j=1}^{t+\bar{c}-1} 1[a_j = \bar{a}] r_j$;
    Update $t \leftarrow t + \bar{c}$ and $\tilde{c}_{\bar{a}} \leftarrow 0$;
  Update $b \leftarrow b+1$;
  Compute Z_b
Return $\hat{a}_\beta = \arg\max_a \hat{\mu}_{\beta,a}$;

Empirical Results



SBC and SPB C-Tracking stop quicker even when constrained to a minimal switching limit ($s = 1$) across the batch horizon.



Conclusion: We presented a formulation to control arm switching frequency in fixed-confidence PE, and showed that it is possible to stop quicker even when constrained to a minimal switching limit, s. Our batched algorithms, SBC and SPB C-Tracking empirically demonstrates this.

References

Mwai, Newton, Milad Malekipirbazari, and Fredrik D. Johansson. *"Batched fixed-confidence pure exploration for bandits with switching constraints."* AISTATS 2025, under review.

WASP WALLENBERG AI AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM