A Best Arm Identification Approach for Stochastic Rising Bandits

A. Montenegro, M. Mussi, F. Trovò, M. Restelli and A. M. Metelli
{alessandro.montenegro, marco.mussi, francesco1.trovo, marcello.restelli, albertomaria.metelli}@polimi.it

Motivation

Several real-world scenarios can be formalized as a **Best-Arm Identification** (BAI) problem in the **Stochastic Rising Bandits** (SRB) setting:

- Combined Algorithm Selection and Hyperparameter Optimization (*CASH*)
- Best Model Selection
- Selection of Athletes for Competitions

Contributions

- Extension of the SRB setting to the **fixed-budget BAI** problem
- Setting **lower bound** on the error probability
- **Two algorithms** solving the problem:
  - `R-UCBE`: an optimistic algorithm
  - `R-SR`: a phase-based algorithm
- **Theoretical analysis** of the error probability upper bounds
- **Numerical validation** on synthetic and real-world data

Setting - Overview

Fixed Budget BAI for SRB

Reward	$x_t = \underbrace{\mu_{I_t}(N_{I_t,t})}_{\text{Expected reward}} + \underbrace{\eta_t}_{\text{Noise}}$
Budget	T
Best Arm	$i^*(T) := \arg\max_{i \in [\![K]\!]} \mu_i(T)$
Growth Rate	$\gamma_i(n) := \mu_i(n+1) - \mu_i(n)$

Goal

Minimize Error Probability

$$e_T(\mathfrak{A}) := \mathbb{P}_{\mathfrak{A}}(\hat{I}^*(T) \neq i^*(T))$$

Assumptions

Rising Bandits

Non-decreasing $\gamma_i(n) \geqslant 0$

Concave $\gamma_i(n+1) \leqslant \gamma_i(n)$

Bounded Growth Rate

$\gamma_i(n) \leqslant c n^{-\beta}$

$c \geqslant 0$ and $\beta > 1$

Setting - Lower Bound

Sub-optimality Gap $\Delta_i(T) := \mu_{i^*(T)}(T) - \mu_i(T)$

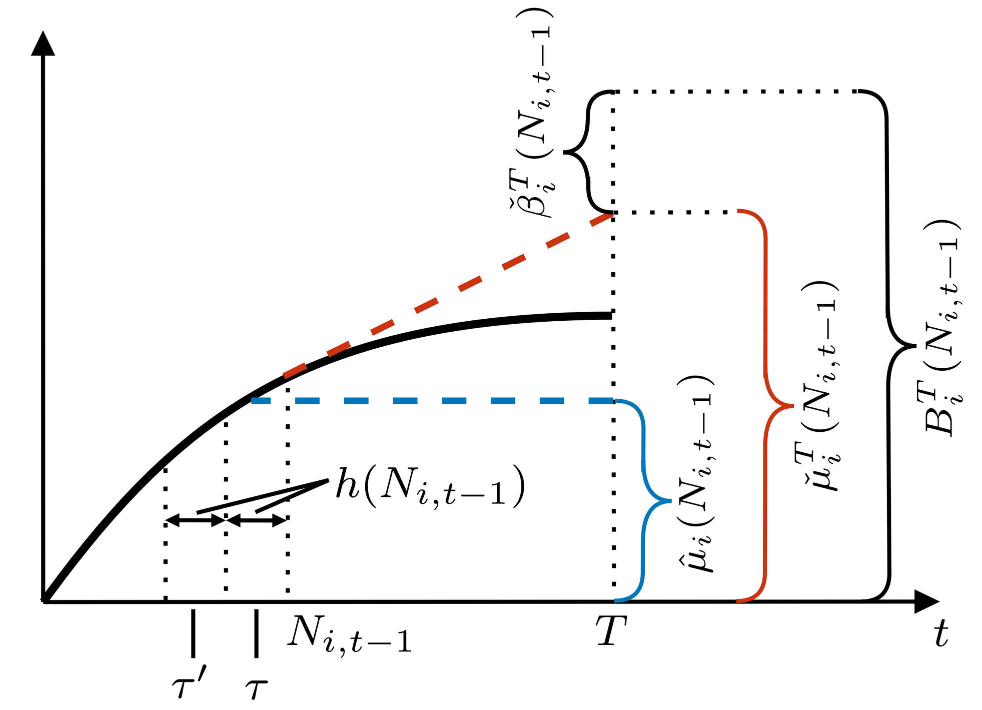
Error Probability

$$e_T(\mathcal{U}) \geqslant \frac{1}{4} \exp\left(-\frac{8T}{\sigma^2 \sum_{i \neq i^*(T)} \frac{1}{\Delta_i(T)^2}}\right)$$

Time Budget

$$T \geqslant \sum_{i \neq i^*(T)} \frac{1}{\Delta_i(T)^{1/(\beta-1)}}$$

Estimators



Pessimistic Estimator

$$\hat{\mu}_i(N_{i,t-1}) := \frac{1}{h(N_{i,t-1})} \sum_{\tau \in \mathcal{T}_{i,t}} x_\tau$$

Optimistic Estimator

$$\check{\mu}_i^T(N_{i,t-1}) := \hat{\mu}_i(N_{i,t-1}) + \sum_{(\tau,\tau') \in \mathcal{S}_{i,t}} (T-j) \frac{x_\tau - x_{\tau'}}{h(N_{i,t-1})^2}$$

Algorithms

Algorithm 1: `R-UCBE`.

Input: Time budget T, Number of arms K, Window size ε, Exploration parameter a
Initialize $N_{i,0} = 0$, $B_i^T(0) = +\infty, \forall i \in [\![K]\!]$
for $t \in [\![T]\!]$ **do**
  Compute $I_t \in \arg\max_{i \in [\![K]\!]} B_i^T(N_{i,t-1})$
  Pull arm I_t and observe x_t
  Update $N_{I_t,t}$
  Update $\check{\mu}_{I_t}^T(N_{I_t,t})$ and $\check{\beta}_{I_t}^T(N_{I_t,t})$
  Compute $B_{I_t}^T(N_{I_t,t}) = \check{\mu}_{I_t}^T(N_{I_t,t}) + \check{\beta}_{I_t}^T(N_{I_t,t})$
end
Recommend $\hat{I}^*(T) \in \arg\max_{i \in [\![K]\!]} B_i^T(N_{i,T})$

Algorithm 2: `R-SR`.

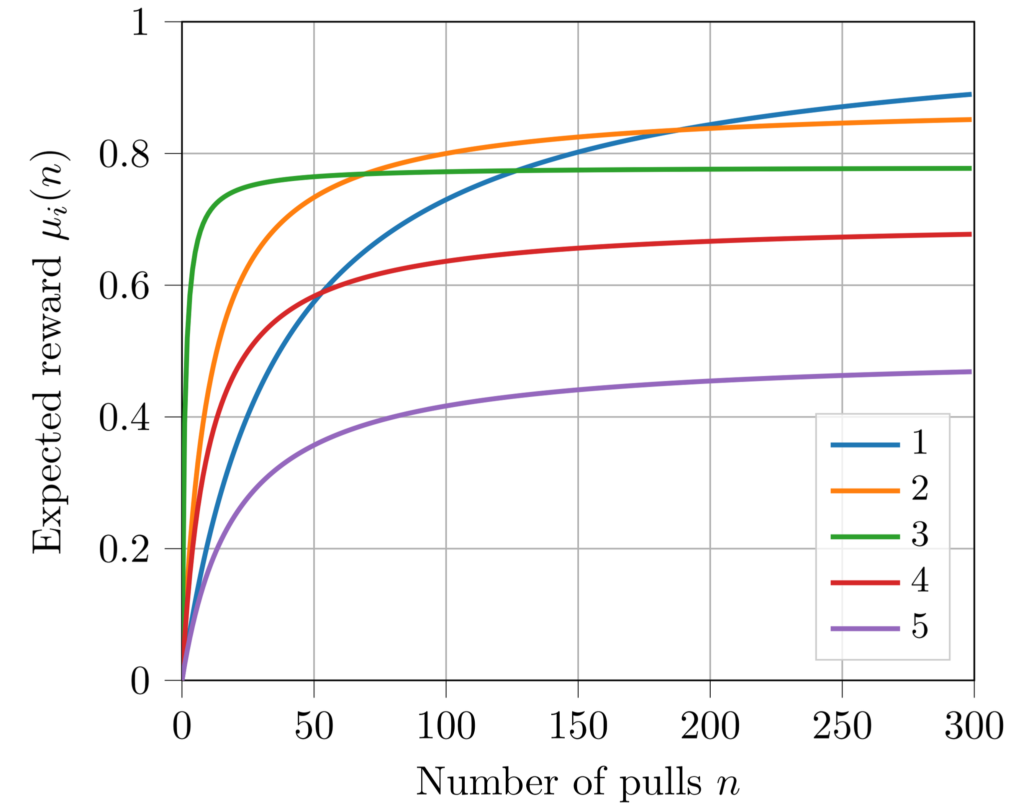
Input: Time budget T, Number of arms K, Window size ε
Initialize $t \leftarrow 1$, $N_0 = 0$, $\mathcal{X}_0 = [\![K]\!]$
for $j \in [\![K-1]\!]$ **do**
  for $i \in \mathcal{X}_{j-1}$ **do**
    Pull $N_j - N_{j-1}$ times
    Update $\hat{\mu}_i(N_j)$
    $t \leftarrow t + N_j - N_{j-1}$
  end
  Define $\overline{I}_j \in \arg\min_{i \in \mathcal{X}_{j-1}} \hat{\mu}_i(N_j)$
  Update $\mathcal{X}_j = \mathcal{X}_{j-1} \setminus \{\overline{I}_j\}$
end
Recommend $\hat{I}^*(T) \in \mathcal{X}_{K-1}$ (unique)

Theoretical Guarantees

	Error Probability $e_T(\cdot)$	Time Budget T
`R-UCBE`	$2\,T\,K\,\exp\left(-\frac{a}{10}\right)$	$\begin{cases} \left(c^{\frac{1}{\beta}}(1-2\varepsilon)^{-1}\left(\sum_{i \neq i^*(T)} \frac{1}{\Delta_i^{1/\beta}(T)}\right) + (K-1)\right)^{\frac{\beta}{\beta-1}} & \text{if } \beta \in (1, 3/2) \\ \left(c^{\frac{2}{3}}(1-2\varepsilon)^{-\frac{2}{3}\beta}\left(\sum_{i \neq i^*(T)} \frac{1}{\Delta_i^{2/3}(T)}\right) + (K-1)\right)^{3} & \text{if } \beta \in [3/2, +\infty) \end{cases}$
`R-SR`	$\frac{K(K-1)}{2} \exp\left(-\frac{\varepsilon}{8\sigma^2} \frac{T-K}{\overline{\log}(K) \max_{i \in [\![K]\!]} \left\{ i \Delta_{(i)}^{-2}(T) \right\}}\right)$	$2^{\frac{1+\beta}{\beta-1}} c^{\frac{1}{\beta-1}} \overline{\log}(K)^{\frac{\beta}{\beta-1}} \max_{i \in [\![2,K]\!]} \left\{ i^{\frac{\beta}{\beta-1}} \Delta_{(i)}(T)^{-\frac{1}{\beta-1}} \right\}$

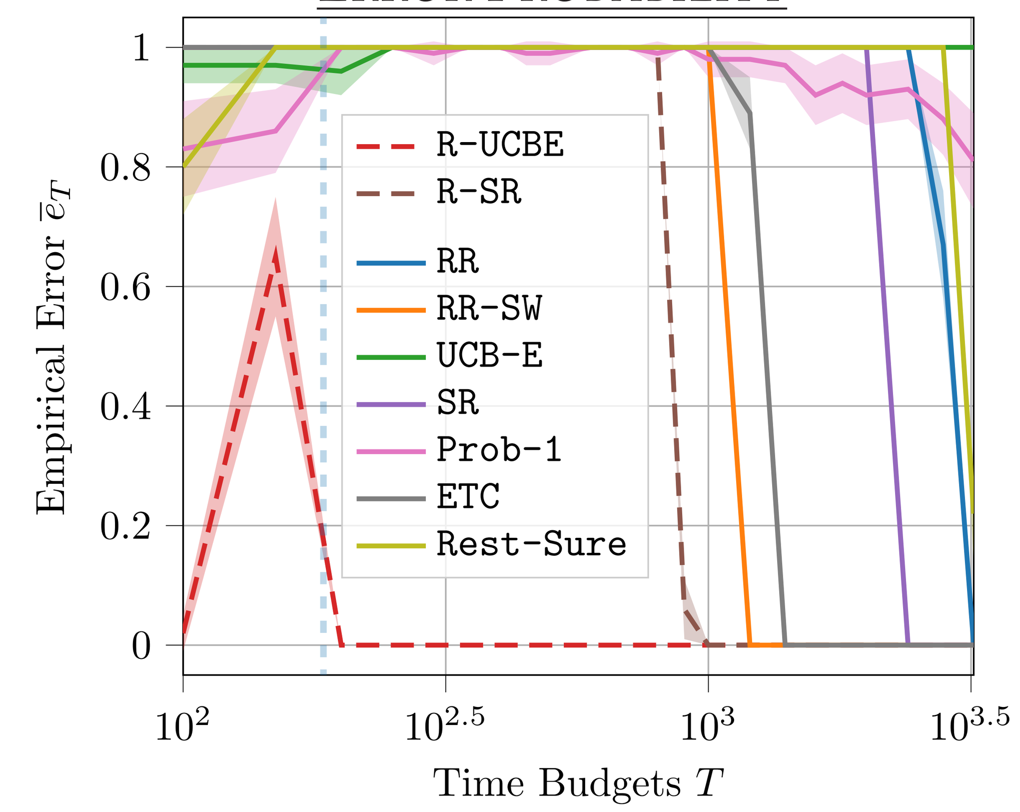
Experimental Validation

Setting

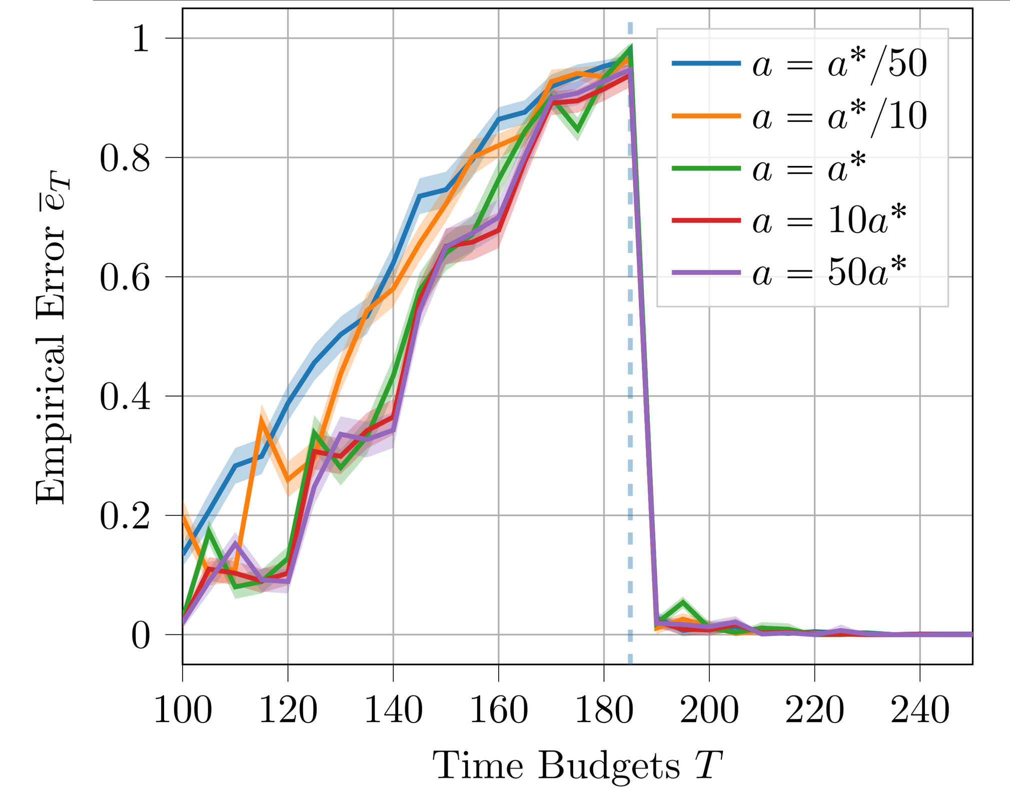


Results

Error Probability



R-UCBE Hyperparameter Sensitivity



References

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *COLT*, 2010.

Alberto Maria Metelli, Francesco Trovò, Matteo Pirola, and Marcello Restelli. Stochastic rising bandits. In *ICML*, 2022.