Best Arm Identification with Resource Constraints

Zitian Li, Wang Chi Cheung

National University of Singapore

INTRODUCTION

Model Setting and Notations

- K: Number of arms.
- L: Types of resources.
- $C = (C_\ell)_{\ell=1}^L \in \mathbb{R}_{>0}^L$: C_ℓ is the Budget of type ℓ resource.
- ν_k: The distribution on the $(L+1)$-variate outcome $(R_k; D_{1,k}, \dots, D_{L,k})$, received by pulling arm k. Assume $\Pr(D_{\ell,k} \in [0,1]) = 1$, and $R_k; D_{1,k}, \dots, D_{L,k}$ can be arbitrarily correlated.
- r_k: Mean reward $\mathbb{E}[R_k] = r_k$ for each $k \in [K]$, $\Delta_k = \max_{a \in [K]} r_a - r_k$.
- $d_{\ell,k}$: Mean consumption $\mathbb{E}[D_{\ell,k}] = d_{\ell,k}$ for each $\ell \in [L], k \in [K]$. Assume $d_{\ell,k} > 0$ for each $\ell \in [L], k \in [K]$.
- $\{d_{\ell,(k)}\}_{k=1}^K$: A permutation of $\{d_{\ell,k}\}_{k=1}^K$, such that $d_{\ell,(1)} \ge d_{\ell,(2)} \ge \dots \ge d_{\ell,(K)}$. This notation is for analysis.
- $A_t \in [K]$: Action in round t.
- $O(t) = (R(t); D_1(t), \dots, D_L(t))$: Received outcomes in round t.
- $H(t) = \{(A_s, O(s))\}_{s=1}^t$: History collected up to time t.
- $I_\ell^{(q)}$: Consumption of resources ℓ in phase q.

Dynamics and Model Uncertainty

At each round $t = 1, 2, \cdots$, the DM

1. Pull an arm $A_t \in [K]$, A_t is $\sigma(H(t-1))$-measurable,
2. Receive the outcome $O(t) = (R(t); D_1(t), \dots, D_L(t)) \sim \nu_{A(t)}$,
3. Update the history $H(t) = H(t-1) \cup \{(A_t, O(t))\}$

The agent stops at the end of time step τ, τ is a stopping time with respect to the filtration $\{\sigma(H(t))\}_{t=1}^\infty$. Upon stopping, the agent outputs arm $\psi \in [K]$ to be the best arm

Known Information

- $K, L, C = (C_\ell)_{\ell=1}^L \in \mathbb{R}_{>0}^L$.

Unknown Information

- ν_k, r_k, $d_{\ell,k}$ for all $\ell \in [L]$, $k \in [K]$

Objective: Minimize Failure Probability

Without loss of generality, assume $r_1 > r_2 \ge \dots \ge r_K$. The agent aims to find a strategy (π, τ, ψ) to maximize $\Pr(\psi = 1)$, subject to

$$\Pr\left(\sum_{t=1}^{\tau} D_\ell(t) \le C_\ell, \forall \ell \in [L]\right) = 1$$

Two settings

- Stochastic: Described above.
- Deterministic: Further assume $\Pr(D_{\ell,k} = d_{\ell,k}) = 1$.
  Comments: If $L = 1$ and $\Pr(D_{1,k} = 1) = 1$ under deterministic setting, it specializes to the fixed budget BAI problem.

Related Work

- **Best Arm Identification:** Fixed Confidence (Even-Dar et al. [2002], Mannor and Tsitsiklis [2004], Audibert and Bubeck [2010], Karnin et al. [2013], Garivier and Kaufmann [2016]), Fixed Budget (Karnin et al. 2013, Carpentier and Locatelli [2016]), Anytime (Audibert and Bubeck [2010], Jun and Nowak [2016]).
  ZITIAN: I think we should make it shorter
- **Simple Regret:** Bubeck et al [2009], Audibert and Bubeck [2010], Zhao et al. [2022].
- **Bandits with Knapsacks (BwK):** Stochastic BwK(Badanidiyuru et al. [2013], Agrawal and Devanur [2014]), adversarial BwK(Immorlica et al. [2019]), non-stationary BwK(Liu et al. [2022])
- **Cost Aware Bayesian Optimization:** Snoek et al. [2012], Poloczek et al. [2017], Swersky et al. [2013], Lee et al. [2020], Luong et al. [2021].

Algorithm and Upper Bound for Failure Probability

Algorithm: SH-RR

Algorithm: Sequential Halving with Resource Rationing

1. Split each resource into $\lceil \log_2 K \rceil$ parts, initialize $\mathsf{Ration}_\ell^{(q)} = \frac{C_\ell}{\lceil \log_2 K \rceil}$ for each $\ell \in [L]$, $q = 1, \cdots, \lceil \log_2 K \rceil$
2. In phase $q = 1, 2, \cdots, \lceil \log_2 K \rceil$, run uniform sampling with budget $\{\mathsf{Ration}_\ell^{(q)} - 1\}_{\ell=1}^L$ on the survival arms
3. Stop phase q if running out any $\{\mathsf{Ration}_\ell^{(q)} - 1\}_{\ell=1}^L$
   - Remove half of the survival arms based on empirical mean reward, from all the collected data.
   - Transfer unused resource budget to the next phase $\mathsf{Ration}_\ell^{(q+1)} = \mathsf{Ration}_\ell^{(q+1)} + (\mathsf{Ration}_\ell^{(q)} - I_\ell^{(q)})^+$.
4. Only one single survive at the end of phase $\lceil \log_2 K \rceil$. Output it as the predicted best arm.

Upper Bound for Deterministic Setting

Consider a BAIwRC instance Q in the deterministic consumption setting. SH-RR has BAI failure probability $\Pr(\psi \neq 1)$ at most

$$\lceil \log_2 K \rceil K \exp\left(-\frac{1}{4\lceil \log_2 K \rceil} \cdot \min_{\ell \in [L]}\{C_\ell / H_{2,\ell}^{\text{det}}(Q)\}\right)$$

where $H_{2,\ell}^{\text{det}}(Q) = \max_{k \in \{2,\dots,K\}} \left\{\frac{\sum_{j=1}^k d_{\ell,(j)}}{\Delta_k^2}\right\}$.

Upper Bound for Stochastic Setting

Consider a BAIwRC instance Q in the stochastic consumption setting. SH-RR has BAI failure probability $\Pr(\psi \neq 1)$ at most

$$7LK(\log_2 K) \exp\left(-\frac{1}{8\lceil \log_2 K \rceil} \cdot \min_{\ell \in [L]}\{C_\ell / H_{2,\ell}^{\text{sto}}(Q)\}\right),$$

where $H_{2,\ell}^{\text{sto}}(Q) = \max_{k \in \{2,\dots,K\}} \left\{\frac{\sum_{j=1}^k f(d_{\ell,(j)})}{\Delta_k^2}\right\}$,

$$f(d) = \begin{cases} e^2 \cdot d & \text{if } d \in [e^{-2}, 1], \\ 2(\log \frac{1}{d})^{-1} & \text{if } d \in (0, e^{-2}). \end{cases}$$

Insights from the Upper Bounds

- Large C_ℓ, smaller $d_{\ell,(k)}$, Larger Δ_k, all lead to small upper bound. min operator implies the bottleneck resource.
- $d_{\ell,(j)}$ in $H_{2,\ell}^{\text{det}}(Q)$, but *effective consumption* $f(d_{\ell,(j)})$ in $H_{2,\ell}^{\text{sto}}(Q)$
  - Non-linear f: the impact of randomness in resource consumption.
  - $d \nearrow \Rightarrow f(d) \nearrow$: higher mean consumption leads to a higher level of usage.
  - $f(d) > d$, and $\lim_{d \to 0} f(d)/d = \infty$, stating that the stochastic setting can be strictly harder than deterministic setting.

Note: from $\max_{k \in \{2,\dots,K\}} \left\{\frac{\sum_{j=1}^k d_{\ell,(j)}}{\Delta_k^2}\right\}$ to $\max_{k \in \{2,\dots,K\}} \left\{\frac{\sum_{j=1}^k d_{\ell,j}}{\Delta_k^2}\right\}$ is impossible.

Lower Bound for Failure Probability

Notations

Requirements of parameters

- (a) Let $\{r_k\}_{k=1}^K$ be any fixed sequence such that $1/2 = r_1 \ge r_2 \ge \cdots \ge r_K \ge 1/4$
- (b) Let $\{\{d_{\ell,(k)}\}_{k=1}^K\}_{\ell \in [L]}$ be any fixed sequence such that $d_{\ell,(1)} \ge d_{\ell,(2)} \ge \cdots \ge d_{\ell,(K)}$ for all $\ell \in [L]$, and $d_{\ell,(k)} \in (0,1]$ for all $\ell \in [L], k \in [K]$

Lower Bound for Deterministic Setting

Define instance $Q^{(i)}$, pulling arm $k \in [K]$ samples

- $R_k \sim \text{Bern}(r_k^{(i)})$, where $r_k^{(i)} = r_k, k \neq i$; $r_k^{(i)} = 1 - r_k, k = i$
- $d_{\ell,k} = d_{\ell,(2)}, k = 1$; $d_{\ell,k} = d_{\ell,(1)}, k = 2$; $d_{\ell,k} = d_{\ell,(k)}$, else

With $\{r_k\}_{k \in [K]}, \{d_{\ell,(k)}\}_{\ell,k}$ being fixed but arbitrary parameters s.t. (a) (b) holds and instances $\{Q^{(i)}\}_{i=1}^K$, when $C_1, \dots, C_L$ are sufficiently large, for any strategy there exists an instance $Q^{(i)}$ (where $i \in [K]$) such that

$$\Pr_i(\psi \neq i) \ge \frac{1}{6} \exp\left(-122 \cdot \min_{\ell \in [L]}\{C_\ell / H_{2,\ell}^{\text{det}}(Q)\}\right),$$

where $\Pr_i(\cdot)$ is the prob measure over the alg and instance $Q^{(i)}$.

Lower Bound for Stochastic Setting

Consider a fixed but arbitrary function $g : [0, +\infty) \to [0, +\infty)$ that is increasing and $\lim_{d \to 0^+} \frac{1}{g(d) \log \frac{1}{d}} = +\infty$, $g(0) = 0$. For any $i \in \{2, \cdots, K\}$, any $\{r_k\}_{k=1}^K$ satisfying (a), any $\{d_{\ell,(k)}^0\}_{k=1,\ell=1}^{K,L}$ satisfying (b), we can find $\bar{c} \in (0,1)$, such that $\forall c \in (0, \bar{c})$ and large enough $\{C_\ell\}_{\ell=1}^L$, by taking $d_{\ell,(j)} = c d_{\ell,(j)}^0, \forall j \in [K], \forall \ell \in [L]$, construct the above $\{Q^{(i)}\}_{i=1}^K$ with Gaussian Reward $N(r_k^{(i)}, 1)$, Bernoulli Consumption and jointly independent R_k, $\{D_{\ell,k}\}_{\ell,k}^{L,K}$, the following performance lower bound holds for any strategy:

$$\max_{j \in \{1,i\}} \Pr_{Q^{(j)}}(\psi \neq j) \ge \exp\left(-2 \min_{\ell \in [L]} \frac{C_\ell}{\tilde{H}_{2,\ell}^{\text{sto}}}\right),$$

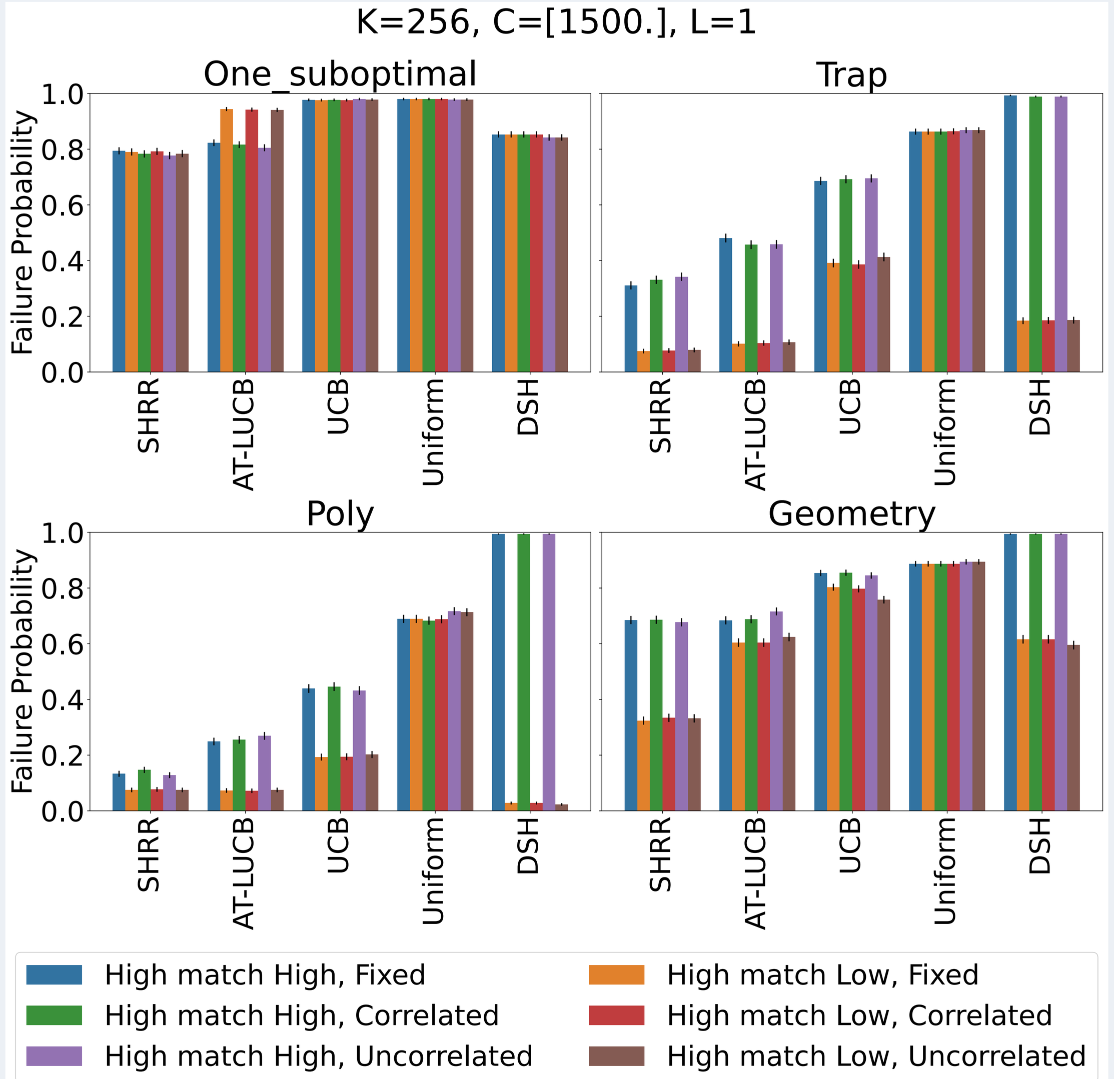
where $\tilde{H}_{2,\ell}^{\text{sto}} = \max_{k \in \{2,3,\cdots,K\}} \left(\sum_{j=1}^k g(d_{\ell,(j)})\right) / \Delta_k^2$.

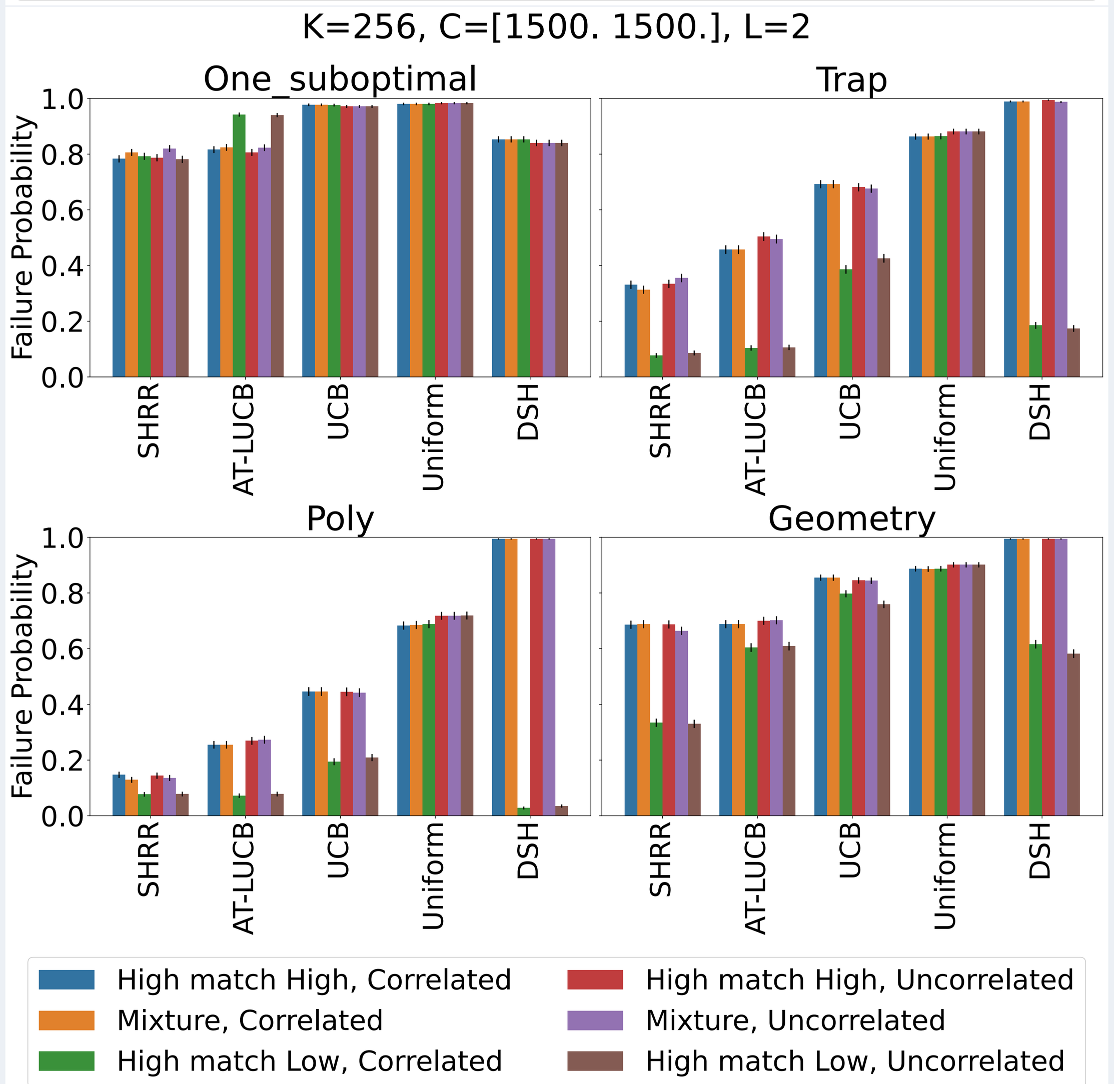
Numeric Experiments

Numeric Experiments

Conduct comparison between SH-RR and ATLUCB(Jun and Nowak [2016]), UCB(Bubeck et al. [2009]), Sequential Halving(Karnin et al. 2013), and Uniform Sampling on different synthesis setting.

- Reward type: 'One Suboptimal', 'Trap', 'Poly', 'Geometry'.
- Matching relationship between Reward and Consumption "High Match High", "High Match Low", "Mix".
- Reward and Consumptions are independent or dependent.





Conclusion: SH-RR remains competitive in both theoretical and numeric performance.