Quantile Multi-Armed Bandits with 1-bit Feedback

Ivan Lau Jonathan Scarlett

Algorithmic Learning Theory 2025

Vanilla Best-Arm Identification

BAI problem (fixed confidence setting):

- K arms with mean rewards $\mu_1, \ldots, \mu_K$
- Learner's goal: Identify the best arm (highest mean) with $\Pr[\text{error}] \leq \delta$ and uses as few samples/arm pulls T as possible
  - $T = O\left(\sum_k \Delta_k^{-2} \log\log \frac{K}{\delta}\right)$, where $\Delta_k = \mu^* - \mu_k$

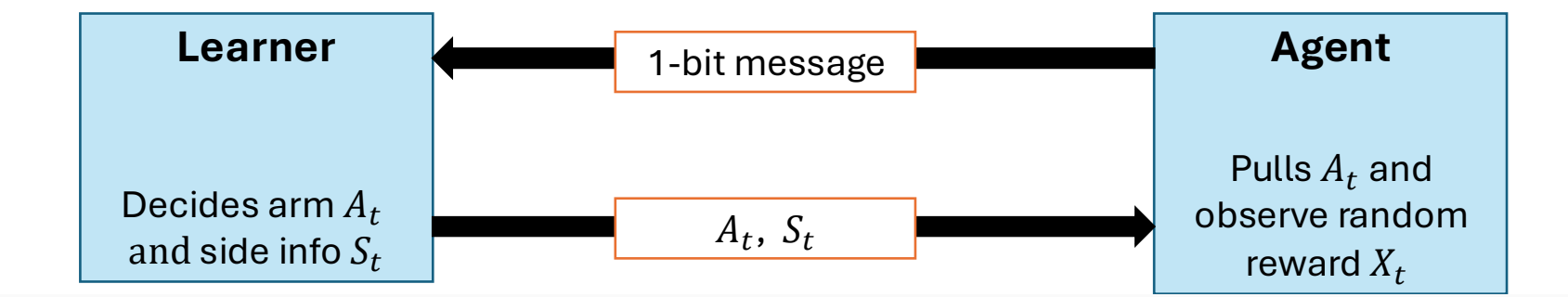
Potential Issues:

- Mean may not always be the ideal performance measure
  - Portfolio 1: $\Pr[\$10^6] = 1$ vs.
    Portfolio 2: $\Pr[\$10^{11}] = 0.01$ and $\Pr[-\$10^9] = 0.99$
- Learner needs precise observations of rewards
  - Uplink communication from the sensor to the server (learner) can be restrictive in some applications

Our BAI: Quantile + 1-bit Feedback Constraint

Problem setup:

- K arms with $F_1^{-1}(q), \ldots, F_K^{-1}(q) \in [0, \lambda]$
  - For simplicity, focus on $q = 0.5$ (median)
- At round $t = 1, \ldots$



- Goal: With probability $\geq 1 - \delta$, find ϵ-optimal arm

$$F_k^{-1}(0.5) \geq \max_a F_a^{-1}(0.5) - \epsilon$$

using as few pulls as possible

Algorithm Outline (I)

Elimination (Even-Dar et al. 2006, Nikolakakis et al. 2021)

- Maintain confidence intervals $\mathrm{LCB} \leq$ median $\leq \mathrm{UCB}$
- Eliminate suboptimal arms

  e.g. $\mathrm{UCB}(\text{arm 1}) \leq \mathrm{LCB}(\text{arm 2}) \implies$ eliminate arm 1

- Terminate when a good arm is found:

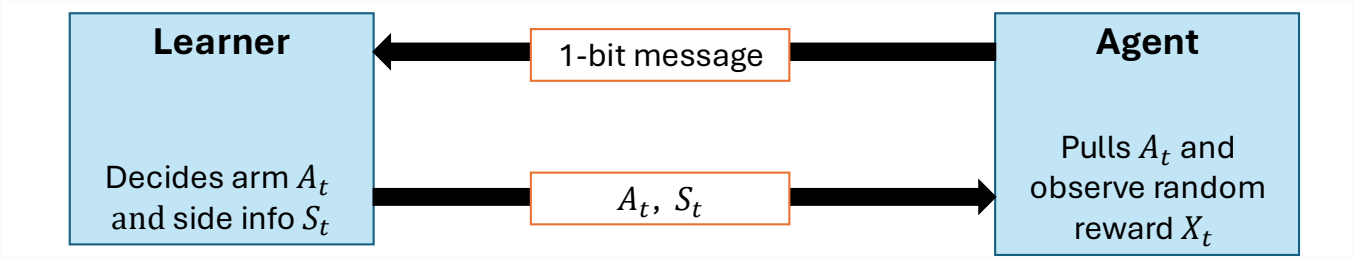
$$\mathrm{LCB}(k) \geq \max_{a \neq k} \mathrm{UCB}(a) - \epsilon$$

How to compute LCB and UCB with no access to empirical median (communication constraint)?

Algorithm Outline (II)

Noisy Binary Search:

- Learner sends 'Is $X_t \leq \gamma_t$?" to the agent as side info S_t
- Based on the 1-bit comparison feedback $\mathbf{1}(X_t \leq \gamma_t)$ from agent, the learner uses **noisy binary search** on $[0, \lambda]$ to compute LCB and UCB
- This costs additional $O(\log(\lambda/\epsilon))$ compared to directly using empirical quantiles (see upper bound later)



Arm Gap (Grossly Simplified)

For non ϵ-optimal arms:

$$\Delta_k := \sup \left\{ \Delta \geq 0 : F_k^{-1}(0.5 + \Delta) \leq \max_a F_a^{-1}(0.5 - \Delta) \right\}$$

For ϵ-optimal arms:

$$\Delta_k := \sup \left\{ \Delta \geq 0 : F_k^{-1}(0.5 - \Delta) \geq \max_{a \neq k} F_a^{-1}(0.5 + \Delta) - \epsilon \right\}$$

Solvability theorem:

- $\max_k \Delta_k > 0 \implies$ instance is solvable by our algorithm
- $\max_k \Delta_k = 0 \implies$ instance is unsolvable (any PAC algorithm can't solve reliably)

Upper and Lower Bounds

Instance Dependent Upper bound:

With probability $\geq 1 - \delta$, our algorithm

- Returns ϵ-optimal arm
- $O\big(\sum_k \Delta_k^{-2}\big[\log(\delta^{-1}) + \log(K\Delta_k^{-1}) + \underbrace{\log(\lambda/\epsilon)}_{\text{noisy binary search}}\ \big]\big)$ samples

Worst-Case Lower bound:

For any (ϵ, δ)-PAC algorithm, there exists an instance requiring

$$\mathbb{E}[T] \geq \Omega\left(\sum_k \Delta_k^{-2} \log(\delta^{-1})\right)$$

Lower bound holds even in the absence of feedback constraints:
1-bit constraint impacts the sample complexity minimally

Conclusion

- Studied a variant of BAI - best quantile, 1-bit feedback
- Algorithm based on elimination + noisy binary search
- Near-matching upper and lower bounds
- 1-bit constraint impacts the sample complexity minimally

References

Even-Dar, E., Mannor, S., Mansour, Y. & Mahadevan, S. (2006), 'Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems.', *Journal of machine learning research* **7**(6).

Nikolakakis, K. E., Kalogerias, D. S., Sheffet, O. & Sarwate, A. D. (2021), 'Quantile multi-armed bandits: Optimal best-arm identification and a differentially private scheme', *IEEE Journal on Selected Areas in Information Theory* **2**(2), 534–548.