Exploration with Limited Memory: Streaming Algorithms for Coin Tossing, Noisy Comparisons, and Multi-Armed Bandits

Sepehr Assadi, Chen Wang
Rutgers University

Abstract

- Finding the most biased coin by tossing -- a classical exploration problem in computer science and machine learning.
- Assuming a gap parameter Δ, elimination-based algorithms have provided solution with $O(\frac{n}{\Delta^2})$ coin tosses which matches the lower bound.
- However, these algorithms inherently require storing all the coins, which is not memory-efficient.
- We studied the sample-space trade-off under the streaming coin tossing mode: the algorithm can only toss an incoming or stored coin.
- We designed an algorithm which only stores **a single extra coin**, which means the sample-space trade-off does not exist.
- En route to the one-coin algorithm, we also proposed preliminary memory-efficient algorithms with $O(\log(n))$, $O(\log\log(n))$ and $O(\log*(n))$ stored coins.
- Extensions of our main algorithm includes finding the k most biased coins and other exploration problems. E.g.. Finding top-k elements using noisy comparisons; Finding an ε-best arm in stochastic multi-armed bandits.

Background

Goal: Find the **most biased coin (denote as coin*)** with **high constant probability** by tossing each coin several times.

Parameters:
n The number of coins;
Δ The gap between the most and second-most biased coins

A good algorithm should:
✓ Return the most biased coin w.h.p. (correctness)
✓ Use small number of tosses (sample complexity)
✓ Store small number of coins (space complexity)

A Naïve Algorithm:
- Toss each coin $O(\frac{\log(n)}{\Delta^2})$ times
- Sample complexity $O(\frac{n}{\Delta^2}\log(n))$
- Space complexity: 1 coin

Median Elimination Algorithm (Even-Dar et al., 2006):
- Round 1: toss each coin $O(\frac{1}{\Delta^2})$ times
- Round 2+: eliminate ½ of the coins; increase 1.5x coin tosses
- Sample complexity $O(\frac{n}{\Delta^2})$
- Space complexity: $\Omega(n)$ coins

Any sample-space trade-off?
-- streaming model: To toss a past coin, must have stored it.

Our Contribution

Main Theorem (Assadi and Wang, 2020)
There exists a streaming algorithm that given n coins arriving in a stream with the gap parameter Δ and confidence parameter δ, finds the most biased coin with probability at least $1-\delta$ using $O(\frac{n}{\Delta^2}\cdot\log(1/\delta))$ coin tosses and a memory of a single coin.

- No sample-space trade-off!
- ❑ Preliminary : $O(\log(n))$, $O(\log\log(n))$ and $O(\log*(n))$ coins memory algorithms.
- ❑ Additional result: Top-k coin exploration with O(k) coin memory.
- ❑ Additional results: Noisy comparisons and Multi-Armed Bandits.

Preliminary Algorithms

The $O(\log(n))$-Coin memory Algorithm:
- Multiple levels: 4-coin memory per level
- Level 1: toss each coin $\frac{30}{\Delta^2}$ times; send the most biased to the level 2.
- Level 2+: increase the number of tosses by 1.5x
- ✓ Correctness: Probability of losing coin* exponentially decreases.
- ✓ Sample complexity: i-th level: $\frac{30}{\Delta^2}\cdot(1.5)^{i-1}\cdot\frac{n}{2^{i-1}}$, overall $O(\frac{n}{\Delta^2})$.
- ✓ Space Complexity: $O(\log(n))$ levels; each level 4 coins.

The $O(\log\log(n))$ and $O(\log*(n))$ Coin Algorithms:
- $O(\log\log(n))$ memory: stopping at the $\log\log(n)$ level
- $O(\log*(n))$ memory: aggressive selections of coins (iterative logarithm factor) and increments of coin tosses (tower factor) (cf. [Agarwal et al., 2017])

Main Algorithm – One Coin Suffices

Idea:
- ❑ Pick only one coin to store, name as *King.*
- ❑ Worst case $\Theta(n)$ coins challenge the *King* -- give the *King* privilege: only be dethroned if lost multiple levels of challenge.
- ❑ Bound the sample complexity: limit the tosses of the *King by* budget.

Algorithm GAME-OF-COINS:
- For each arriving coin give the *King* a budget of $O(\frac{1}{\Delta^2})$.
- To challenge the *King*, toss both coins $\frac{30}{\Delta^2}\cdot(3)^{i-1}$ times at level i;
- A *King* is defeated *only if* it exhausts all its budget.

Analysis:
- ✓ Sample Complexity: At most $2n\cdot O(\frac{1}{\Delta^2})$ budgets $\rightarrow O(\frac{n}{\Delta^2})$ coin tosses.
- ✓ Space Complexity: Only store 1 coin.
- ✓ Correctness:
  1. The coin* can exhaust the budget of other *King* (soundness)
  2. If coin* as the *King* $\rightarrow$ budget sufficient *in expectation.*
  3. Control the variance:
     a) The budget behaves like random walks (but with flexible length).
     b) The challenging rule $\rightarrow$ budget distribution sub-exponential.
     c) Beating the union bound by Bernstein inequality (completeness).

Extensions of the Algorithm

Algorithm for top-k coins:
- Main technical contribution -- a delayed challenging rule & a potential function argument.
- Avoid eliminating any top-k coin -- use a buffer to swap defeated coins (correctness).
- Number of coins eventually decreases -- bounded sample complexity.

Noisy Comparisons and ε-PAC Multi-Armed Bandit (MAB):
- Noisy comparison – O(k) space algorithm for finding top-k elements.
- No gap guarantee -- a $O(\log*(n))$ space algorithm. Most recently, an extension to a 2-arm algorithm.

Extensions and open problems:
- The instance-sensitive sample complexity: $H_2 := O(\sum_{i>1}\frac{1}{\Delta_i^2}\log\log(\frac{1}{\Delta_i}))$.
- Single-pass: achievable with random arrival of coins and a value $O(H_2)$.
- Single-pass with lower bounds; arbitrary stream with $O(\log(\frac{1}{\Delta_2}))$ passes [Jin et al., 2021].
- *Open*: tight number of passes to achieve $O(H_2)$ sample complexity.

References

Assadi S, Wang C. Exploration with Limited Memory: Streaming Algorithms for Coin Tossing, Noisy Comparisons, and Multi-Armed Bandits. The 52nd Annual ACM Symposium on Theory of Computing (STOC) 2020

Even-Dar E, Mannor S, Mansour Y. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. Journal of Machine Learning Research (JMLR). 2006

Agarwal A, Agarwal S, Assadi S, Khanna S. Learning with limited rounds of adaptivity: Coin tossing, multi-armed bandits, and ranking from pairwise comparisons. In Conference on Learning Theory (COLT) 2017 Jun 18

Jin T, Huang K, Tang J, Xiao X. Optimal Streaming Algorithms for Multi-Armed Bandits. In International Conference on Machine Learning (ICML) 2021 Jul 1