Distribution Oblivious, Risk-Aware Algorithms for Multi-Armed Bandits with Unbounded Rewards

Anmol Kagrecha[1], Jayakrishnan Nair[1], Krishna Jagannathan[2]
Department of Electrical Engineering, [1]IIT Bombay, [2]IIT Madras

Motivation

Distribution Obliviousness

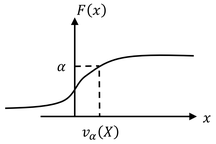
- Reward distributions in many Multi-Armed Bandit (MAB) problems are assumed to have known and bounded support.
- When unbounded rewards are considered, bounds on moments/tails are assumed to be known & are used to construct MAB algorithms.
- Violates the spirit of online learning, motivates distribution obliviousness.

Risk Awareness

- In many MAB problems, goodness of an arm is quantified using the expected return, which is a risk-neutral metric.
- In some applications, particularly in finance, it is of interest to balance the expected return and the risk associated with an arm.

Capturing Risk

- Consider random variable X depicting loss and a given confidence level $\alpha \in (0,1)$
- Worst loss at confidence α is Value at Risk (VaR): $v_\alpha(X) = F_X^{-1}(\alpha)$
- Conditional Value at Risk (CVaR): $c_\alpha(X) = E[X|X \geq v_\alpha(X)]$
- CVaR is a coherent risk unlike VaR; used extensively in portfolio optimization, insurance, etc



Problem Setup

- There are K arms and each pull of arm k yields IID loss $X(k)$
- Assumption: there exists $\varepsilon \in (0,1), B > 0$ such that $E[|X(k)|^{1+\varepsilon}] < B$ for all arms $k \in [K]$
- Algorithm doesn't know ε or B and has to identify the arm k^* minimizing $\xi_1\, \mu[k] + \xi_2\, c_\alpha(k)$ using T pulls
- Performance metric: p_e = Probability(output $\neq k^*$)

Summary of our Results

- Lower bound: for any algorithm, there is an instance ν such that $p_e(\nu) \geq c \exp(-dT\ - o(T))$
- There are non-oblivious algorithms (know ε and B) with $p_e \leq c' \exp(-d'T)$
- Distribution oblivious algorithms using empirical estimators have $p_e \leq c^\dagger T^{-\varepsilon} + o(T^{-\varepsilon})$ and the bound is tight!
- We construct distribution oblivious algorithms with $p_e \leq C\exp(-DT^{1-q})$ for $T > f(q, \text{problem instance})$ where $q \in (0,1)$

Distribution Oblivious Algorithms

Considering only CVaR minimization: $(\xi_1, \xi_2) = (0,1)$; general setting discussed in the paper.

Truncated Empirical CVaR (TEC)

$\{X_i\}_{i=1}^n$: n IID samples of a random variable X;

$X_i^{(b)} = \min(\max(-b, X_i), b), b > 0$

$\{X_{[i]}^{(b)}\}_{i=1}^n$: order statistics such that $X_{[1]}^{(b)} \geq \cdots \geq X_{[n]}^{(b)}$

$$\hat{c}_{n,\alpha}^{(b)}(X) = \hat{c}_{n,\alpha}\left(X^{(b)}\right) = X_{[\lfloor n(1-\alpha) \rfloor]}^{(b)} + \frac{1}{n(1-\alpha)} \sum_{i=1}^{\lfloor n(1-\alpha) \rfloor} X_{[i]}^{(b)} - X_{[\lfloor n(1-\alpha) \rfloor]}^{(b)}$$

Concentration Inequality

Given $\Delta > 0$, $\mathrm{P}\left(\left|c_\alpha(X) - \hat{c}_{n,\alpha}^{(b)}(X)\right| \geq \Delta\right) \leq 6\exp\left(-n(1-\alpha)\frac{\Delta^2}{154b^2}\right)$

for $b > b^* = \max\left(\frac{\Delta}{2}, |v_\alpha(X)|, \left\lceil \frac{2B}{\Delta(1-\alpha)} \right\rceil^{\frac{1}{\varepsilon}}\right)$

b not known to algorithm:*
grow truncation parameter as n^q, $q \in (0,0.5)$

Algorithms

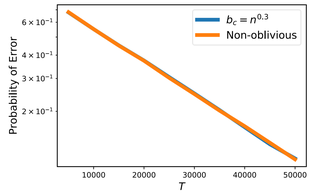
Perform uniform exploration with TEC

Output arm with minimum estimator value

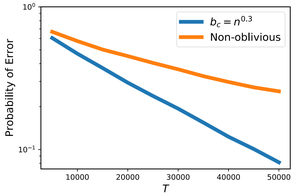
Analysis of algorithms like Successive Rejects discussed in the paper

Theorem: $p_e \leq Cexp(-DT^{1-2q})$ for $T > T^*, q \in (0,0.5)$
where T^* depends on q and the problem instance

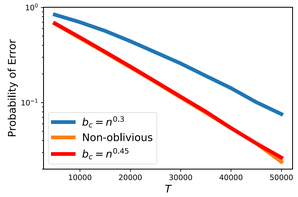
Numerical experiments



Light Tailed Arms



Heavy Tailed Arms



Mixture of LT and HT arms:
Light Tailed Arm Optimal

Discussion

- Numerical experiments suggest that oblivious algorithms perform competitively when compared to non-oblivious algorithms.
- This suggests a need of tighter upper bounds and fundamental lower bounds for distribution oblivious algorithms.
- Interesting to explore distribution oblivious algorithms in the regret minimization and PAC framework.