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

Anmol Kagrecha[1], Jayakrishnan Nair[1] and Krishna Jagannathan[2]

Department of Electrical Engineering, [1]IIT Bombay, [2]IIT Madras

akagrecha@gmail.com, jayakrishnan.nair@ee.iitb.ac.in, krishnaj@ee.iitm.ac.in

Motivation

Distribution Obliviousness

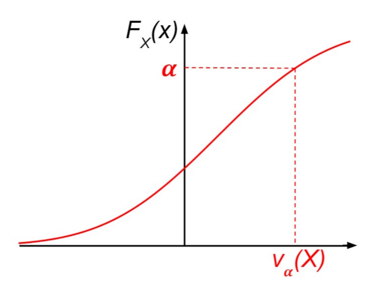
- Reward distributions in many MAB problems are assumed to have known & bounded support.
- For unbounded rewards, moment bounds assumed to be known & used to devise MAB algorithms.
- This violates the spirit of online learning and motivates distribution obliviousness.

Risk Awarness

- In classical MAB problems, goodness of arm is measured by expected return, a risk-neutral metric.
- In applications like finance one needs to balance expected return & risk associated with an arm.

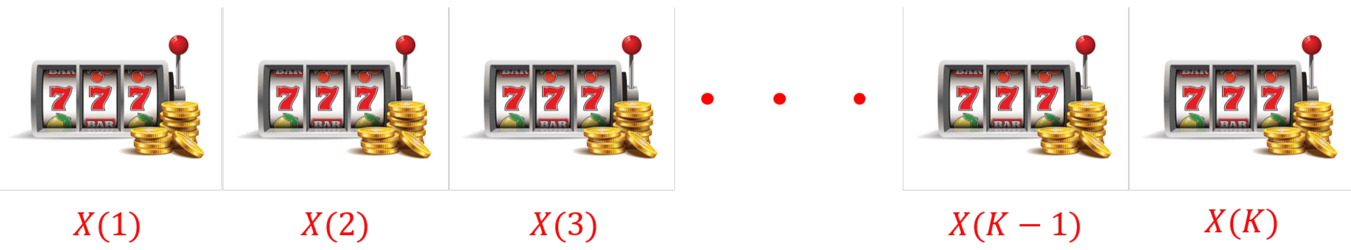
Capturing Risk

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



Depiction of VaR

Problem Setup



- **Assumption**: There exists $\varepsilon \in (0, 1]$, $B > 0$ such that $\mathbb{E}\left[|X(k)|^{1+\varepsilon}\right] < B$ for all $k \in [K]$.
  Allows the arms to be unbounded and even heavy tailed.
  The algorithm doesn't know ε or B.
- **Objective**: Identify arm k^* minimizing $\xi_1\mu(k) + \xi_2 c_\alpha(k), \xi_1, \xi_2 \geq 0$ using T pulls.
  Linear combination of mean and CVaR
- **Performance metric**: Probability of incorrect identification of k^*: $p_e = \Pr(\text{output} \neq k^*)$.

Summary of Results

- **Non-oblivious algorithms** which know ε, B and Δ[2], have $p_e \leq c' \exp(-d'T)$.
- **Lower bound***: Any distribution oblivious consistent algorithm can not have an exponential decay in T for all instances which have some $(1+\varepsilon)^{th}$ moment unbounded.
- **Naive algorithms*** which use empirical estimators have $p_e \leq c^\dagger T^{-\varepsilon}$ and the bound is tight!
  Bounds on p_e decay polynomially instead of exponentially!
- We **construct oblivious algorithms** with $p_e \leq C(q)\exp(-DT^{1-q})$, $q \in (0, 1)$
  Decay in upper bound can be made arbitrarily close to exponential but not exactly equal. As q goes to zero, $C(q)$ goes to infinity.

*: Recent result, not in the paper

Distribution Oblivious Algorithms

Considering only CVaR minimization: $(\xi_1, \xi_2) = (0, 1)$ and Uniform Exploration algorithm.
General linear combinations of CVaR & mean, analysis of Successive Rejects discussed in the paper.

Empirical CVaR

- $\{X_i\}_{i=1}^n$: n IID samples of a random variable X, $\{X_{[i]}\}_{i=1}^n$: order statistics such that $X_{[1]} \geq \cdots \geq X_{[n]}$
- Empirical CVaR estimator is given by

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

- Can be shown that $P(|\hat{c}_{n,\alpha}(X) - c_\alpha(X)| > \Delta) \leq g(\Delta)/n^\varepsilon$ and the inequality is tight.

High variability of heavy tailed arms leads to poor concentration results for empirical estimator!

Truncated Empirical CVaR (TEC)

- Let $X_i^{(b)} = \min(\max(-b, X_i), b)$, $\quad b > 0$ and $\{X_{[i]}^{(b)}\}_{i=1}^n$ be the order statistics of $X_i^{(b)}$.
- TEC is the empirical CVaR of $X^{(b)}$ and is defined as:

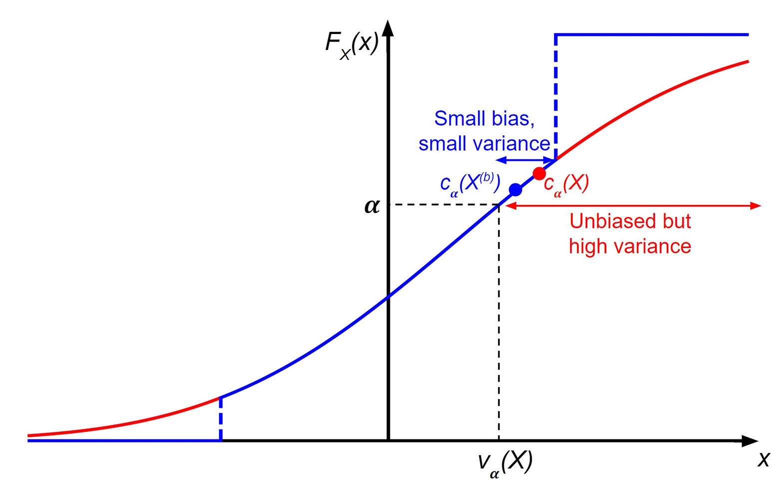
$$\hat{c}_{n,\alpha}^{(b)}(X) = \hat{c}_{n,\alpha}(X^{(b)}) = X^{(b)}_{[\lceil n(1-\alpha) \rceil]} + \frac{1}{n(1-\alpha)}\sum_{i=1}^{\lfloor n(1-\alpha) \rfloor}(X^{(b)}_{[i]} - X^{(b)}_{[\lceil n(1-\alpha) \rceil]})$$

- We prove the following concentration inequality for TEC in the paper:

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

$$\text{for } b > b^* = \max\left(\frac{\Delta}{2}, |v_\alpha(X)|, \left[\frac{2B}{\Delta(1-\alpha)}\right]^{\frac{1}{p-1}}\right)$$

Fixing $b > b^*$ controls variability & ensures bias is at most $\Delta/2$.



Truncation helps to control the bias-variability trade-off!

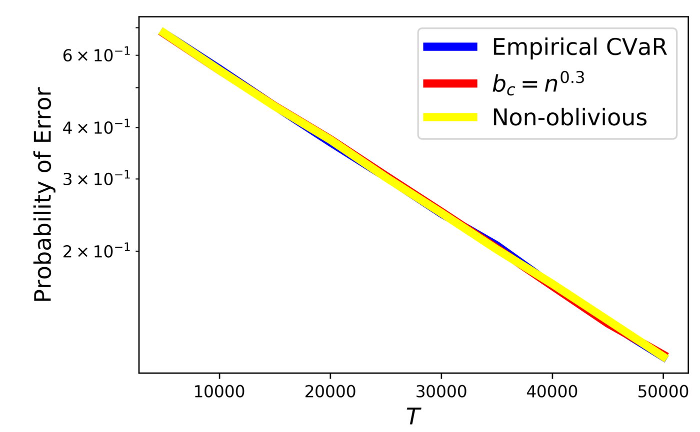
Performance Analysis

- Distribution oblivious UE doesn't know b^* so grow b as $(T/K)^q$ where $q \in (0, 1)$.
- Output the arm which has minimum value for the estimator.
- **Theorem**: $p_e \leq C\exp(-DT^{1-2q})$ *for* $T > T^*$, $q \in (0, 0.5)$ *where* T^* *depends on problem instance and* q.

Numerical Experiments

Successive Rejects is used for all the experiments below. The confidence parameter $\alpha = 0.95$.

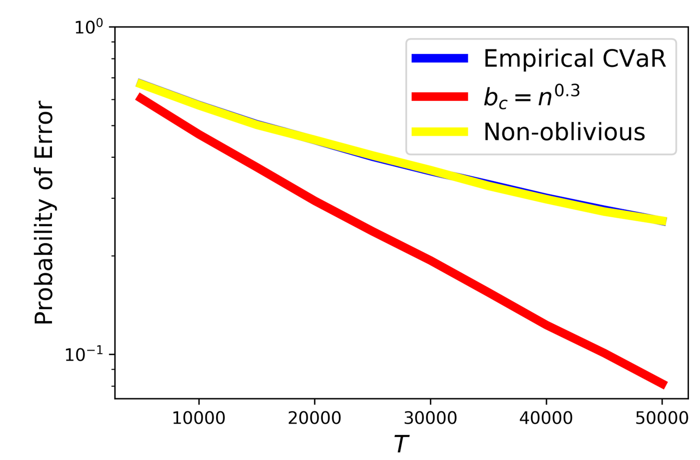
Light Tailed Arms



- Consider 10 arms which are exponentially distributed.
- Optimal arm has a CVaR=2.85 & other arms have CVaR=3.0.

All the algorithms perform equally well

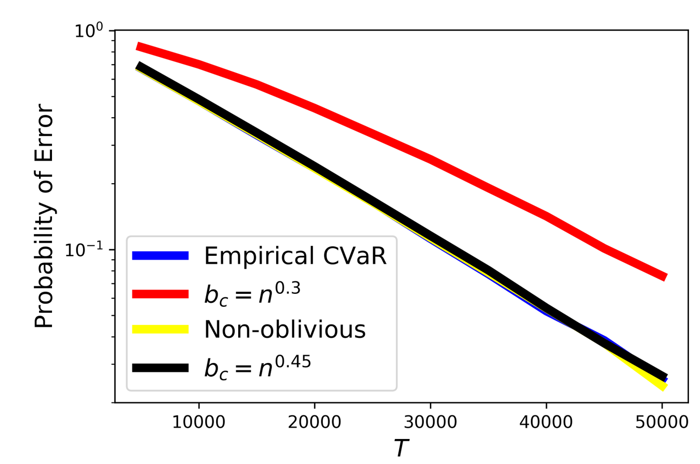
Heavy Tailed Arms



- Consider 10 arms distributed according to Lomax distribution (shape parameter = 2.0).
- Optimal arm has a CVaR=2.55 & other arms have CVaR=3.0.

Growing truncation parameter slowly helps to control the variability of arms!

Mixture of LT and HT arms



- Consider 10 arms, where 5 arms are Exponential and 5 arms are Lomax (shape parameter = 2.0).
- Otimal arm is Exponential and has CVaR=2.55; other arms have CVaR=3.0.
- Truncation leads to greater underestimation of CVaR of HT arms compared to LT arms.

Growing truncation fast or using empirical estimator is beneficial!

Future Directions

- Constructing distribution oblivious algorithms that perform better numerically & are data-driven.
- Constructing distribution oblivious algorithms for fixed confidence and regret minimization settings.

References

[1] Sébastien Bubeck, Nicolo Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013.

[2] Ravi Kumar Kolla, LA Prashanth, Sanjay P Bhat, and Krishna Jagannathan. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters*, 47(1):16–20, 2019.

[3] Xiaotian Yu, Han Shao, Michael R Lyu, and Irwin King. Pure exploration of multi-armed bandits with heavy-tailed payoffs. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 937–946, 2018.