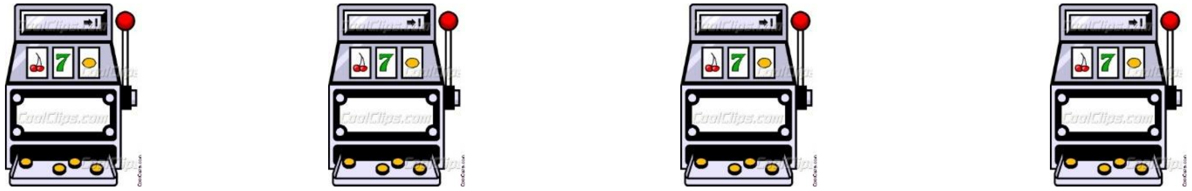
Environment oblivious, risk-aware multi-armed banditry

Jayakrishnan Nair (IITB)

joint work with Anmol Kagrecha (IITB) & Krishna Jagannathan (IITM)

Multi-armed bandit problem

Fundamental problem in online learning: Identify the best among a basket of options



F_1 F_2 F_3 F_4 ← unknown reward distributions

Example: Identify option (arm) with highest mean reward

F_1

F_2

F_3

F_4

Classical setup:

- Rewards have known and bounded support, say $[0,1]$
- Want to identify arm with highest mean reward

Q: What if rewards have unknown/unbounded support (e.g., heavy-tailed)?

A: Limited literature; typically assumes that certain bounds on the moments/tails are known.

Violates spirit of online learning?

Motivates environment oblivious algorithms

F_1

F_2

F_3

F_4

Classical setup:

- Rewards have known and bounded support, say $[0,1]$
- Want to identify arm with highest mean reward

Q: What if I want to be risk-aware in my arm selection?

A: Few results on risk-aware arm selection, none allowing for heavy-tailed rewards

Agenda

- Design environment oblivious MAB algorithms
  - No restrictive assumptions on reward distributions, allow for unbounded support, heavy tails
  - Provable performance guarantees

- Incorporate risk measures in arm selection criterion

This talk: Our first steps in this direction

Preliminaries: Heavy tails

Random variable X is *heavy-tailed* if

$$\limsup_{x\to\infty} \frac{P\left(X > x\right)}{e^{-\nu x}} = \infty \quad \forall \nu > 0$$

Tail is asymptotically `heavier' than exponential

E.g.: Pareto distribution: $P\left(X > x\right) = cx^{-\alpha}$, $\alpha > 0$

Weibull distribution: $P\left(X > x\right) = e^{-cx^{\theta}}$, $\theta \in (0, 1)$

Preliminaries: Heavy tails

Random variable X is *heavy-tailed* if

$$\limsup_{x \to \infty} \frac{P(X > x)}{e^{-\nu x}} = \infty \quad \forall \nu > 0$$

Tail is asymptotically `heavier' than exponential

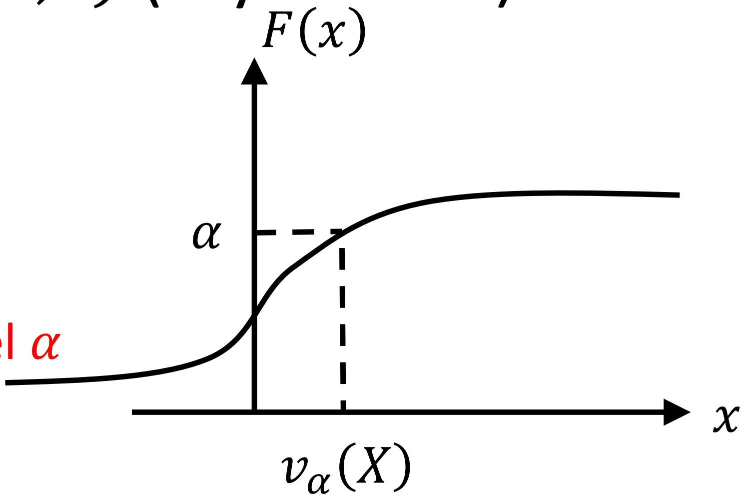
Heavy tails are ubiquitous: incomes, city sizes, Internet file sizes, insurance claims, …
Very little MAB literature allowing heavy tails, none that is environment oblivious

Preliminaries: Capturing risk

For random variable X & confidence level $\alpha \in (0,1)$ (say α=0.95):

Value at Risk (VaR) $v_\alpha(X) := \mathrm{F}^{-1}(\alpha)$

worst case loss corresponding to confidence level α



Preliminaries: Capturing risk

Conditional Value at Risk (CVaR) $c_\alpha(X) \coloneqq E[X|X \geq v_\alpha]$

$$= v_\alpha + \frac{1}{1-\alpha} E[\,X - v_\alpha]^+$$

Expected loss conditioned on `bad event' that loss exceeds VaR

- CVaR is a *coherent* risk measure (unlike VaR)
- Used extensively in portfolio optimization, credit risk assessment, insurance

Model

- K arms
- Each pull yields i.i.d cost/loss $\sim X(i)$

- Assumption: For all arms, $E\left[|X(i)|^{1+\delta_i}\right] < \infty$ for some $\delta_i > 0$

  $\Rightarrow \exists \epsilon \in (0,1), B > 0$ s.t. $E[|X(i)|^{1+\epsilon}] < B$ for all i

  - Only mildly more restrictive that well-posedness
  - Allows for heavy-tailed distributions
  - Algorithm does not know ϵ, B

Model

- K arms
- Each pull yields i.i.d cost/loss $\sim X(i)$
- Assumption: For all arms, $E\left[|X(i)|^{1+\delta_i}\right] < \infty$ for some $\delta_i > 0$

  $\Rightarrow \exists \epsilon \in (0,1), B > 0$ s.t. $E[|X(i)|^{1+\epsilon}] < B$ for all i

- Goal: Identify arm that minimizes $\xi_1 E[X(i)] + \xi_2 c_\alpha(X(i))$ given $\xi_1, \xi_2 \geq 0$

  using T pulls

Pure exploration

This talk: $\xi_1 = 0, \xi_2 = 1$ (CVaR minimization)

Performance metric & fundamental limits

- Performance metric: $p_e = Prob(incorrect\ identification)$
- Lower bound: For any algorithm, $p_e \geq d\, e^{-cT} \quad (c, d > 0)$
- Can design non-oblivious algorithm with $p_e \leq \tilde{d}\, e^{-\tilde{c}T}$

Knows ϵ, B and lower bound on sub-optimality gap

Q: Can oblivious algorithms achieve exponential decay of p_e?

(Naïve) Approach: Empirical estimators

- Perform *uniform exploration*, i.e., pull the arms round robin
- Compute empirical CVaR estimate for each arm

Given i.i.d. observations $X_1, X_2, \dots, X_n$

Let $(X_{[1]}, X_{[2]}, \dots, X_{[n]})$ denote the order statistics, i.e.,

$$X_{[1]} \geq X_{[2]} \geq \cdots \geq X_{[n]}$$

$$\hat{c} = \underbrace{X_{[\lceil n(1-\alpha) \rceil]}}_{\text{VaR estimate}} + \frac{1}{n(1-\alpha)} \sum_{i=1}^{\lfloor n(1-\alpha) \rfloor} X_{[i]} - X_{[\lceil n(1-\alpha) \rceil]}$$

CVaR estimate

(Naïve) Approach: Empirical estimators

- Perform *uniform exploration*, i.e., pull the arms round robin
- Compute empirical CVaR estimate for each arm
- Select arm that minimizes $\hat{c}_i$

Theorem: $p_e \leq \frac{C}{T^\epsilon} + o\left(\frac{1}{T^\epsilon}\right)$.

- Probability of error decays far slower than exponential!
- *This bound is tight.*

(Naïve) Approach: Empirical estimators

Theorem: $p_e \leq \frac{C}{T^\epsilon} + o\left(\frac{1}{T^\epsilon}\right)$.

$$\Uparrow$$

Theorem: With n samples, the empirical CVaR estimator satisfies

$$P(|c_\alpha - \hat{c}| > \Delta) \leq \frac{g(\epsilon, \Delta)}{n^\epsilon} + o\left(\frac{1}{n^\epsilon}\right)$$

bound is tight

- Similar to the concentration inequality for empirical mean
- Empirical estimators highly variable for heavy-tailed distributions

Truncation based approach

Given i.i.d. observations $X_1, X_2, \dots, X_n$

Truncated empirical estimator $\hat{c}^b$ is the estimator corresponding to $X_1^b, X_2^b, \dots, X_n^b$,
where $\mathrm{X_i^b} = (min(max(X_i), -b), b)$

projection of X_i onto $[-b, b]$

- Enables *bias-variance* tradeoff
- Large b implies small bias but high variance
- Small b implies large bias but small variance

Truncation based approach

Theorem: Given $\Delta > 0$,

$$P\left(|c_\alpha(X) - \hat{c}_t(b)| \geq \Delta\right) \leq 6\exp\left(-n(1-\alpha)\frac{\Delta^2}{48b^2}\right)$$

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

ensures bias is at most $\Delta/2$

- But $\bar{b}$ is not known to the algorithm!
- Idea: grow truncation parameter b with the number of pulls
- $b = n^q$ for $q \in (0, 1/2)$ implies, for large enough n,

$$P\left(|c_\alpha(X) - \hat{c}_t(b)| \geq \Delta\right) \leq 6\exp\left(-n^{1-2q}(1-\alpha)\frac{\Delta^2}{48}\right)$$

Truncation based approach

- Perform *uniform exploration*, i.e., pull the arms round robin
- Compute empirical CVaR estimate for each arm, using truncation parameter $b = T^q$, for $q \in (0, {}^1/_2)$
- Select arm that minimizes $\hat{c}_i^b$

$$\text{Theorem: } p_e \leq C\exp\left(-DT^{1-2q}\right) \text{ for } T > T^*,$$
$$\text{where } T^* \text{ depends on the problem instance and } q.$$

- Much stronger guarantee than with empirical estimator
- But probability of error decays slower than exponentially
- Guarantees kick in only for large enough T
- Can be extended to successive rejects

Median-of-bins approach

Given i.i.d. observations $X_1, X_2, \ldots, X_n$

Partition the data into $^n/_k$ bins, each containing k samples
$\hat{c}_j$=Empirical CVaR estimator corresponding to bin j

$$\hat{c}^{mb} = median\left(\hat{c}_1, \hat{c}_1, \cdots, \hat{c}_{n/k}\right)$$

robust to outliers in the data

Theorem: For $k \geq \bar{k}$, where $\bar{k}$ depends on Δ and the dist. of X,

$$P\left(\left|c_\alpha(X) - \hat{c}^{mb}\right| > \Delta\right) \leq e^{-n/8k}.$$

- In MAB setting, we don't know $\bar{k}$
- But can grow k with the number of samples, say $k = T^q$ for $q \in (0,1)$

Median-of-bins approach

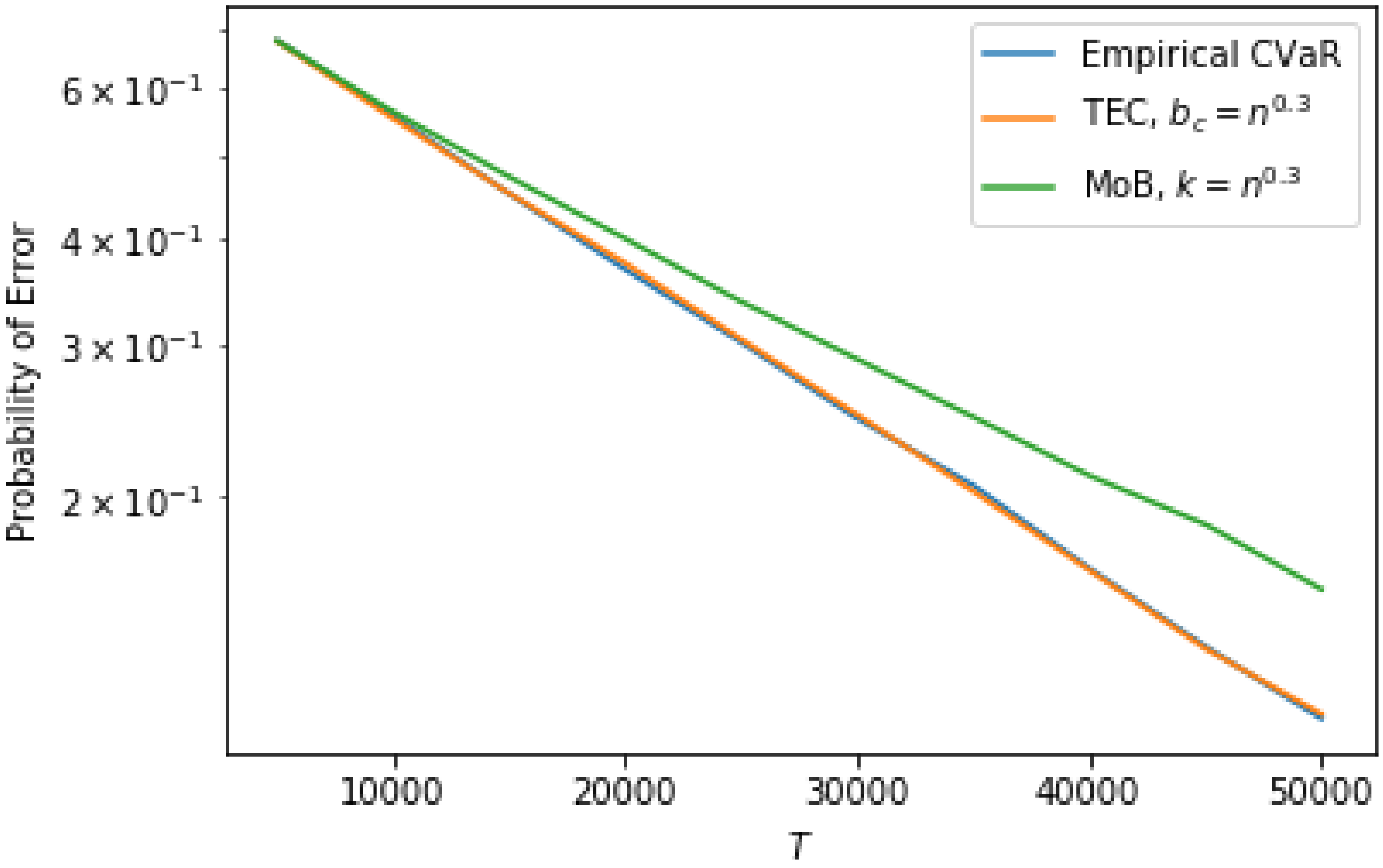
- Perform *uniform exploration*, i.e., pull the arms round robin
- For each arm i, compute *mb* estimator $\hat{c}_i^{mb}$, with each bin having T^q samples, for $q \in (0,1)$
- Select arm that minimizes $\hat{c}_i^{mb}$

Theorem: $p_e \leq C\exp\Big(-DT^{1-q}\Big)$ for $T > T^*$, where T^* depends on the problem instance and q.

- Much stronger guarantee than with empirical estimator
- But probability of error decays slower than exponentially
- Guarantees kick in only for large enough T
- Can be extended to successive rejects

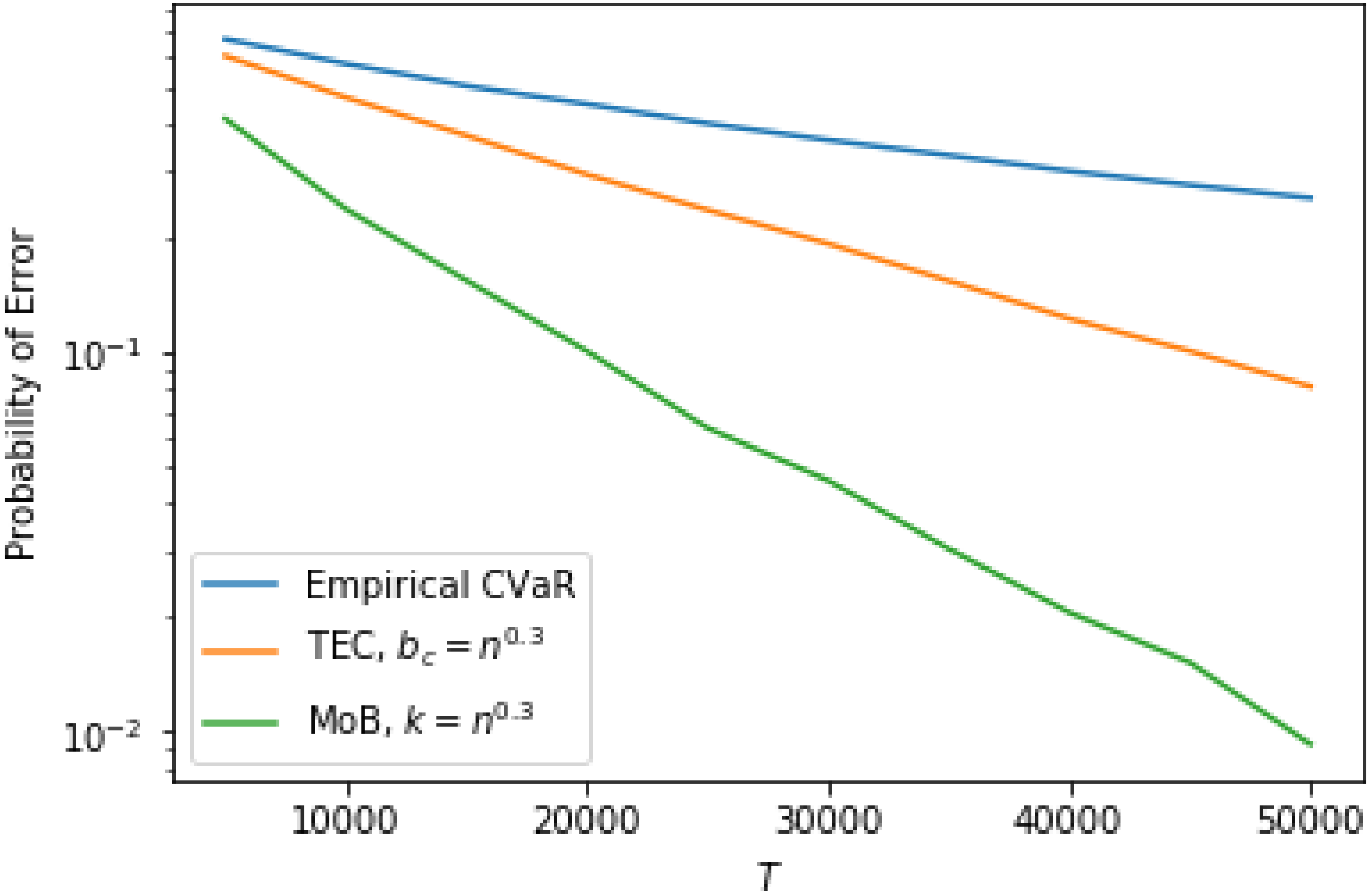
Simulation results

Light-tailed example



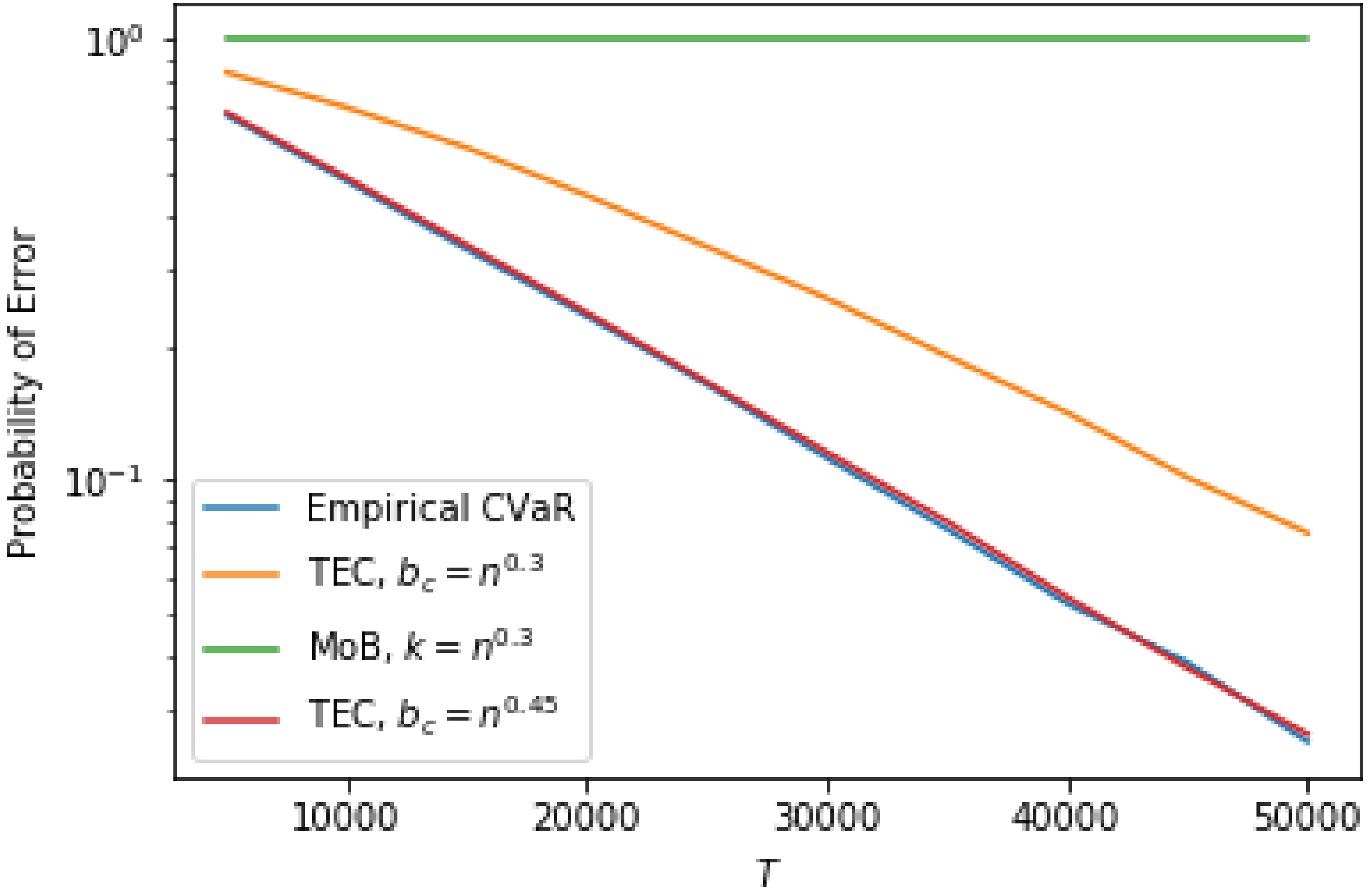
10 arms, exponential loss
Opt. arm: CVaR = 2.85
Rest: CVaR = 3
$\alpha = 0.95$

Heavy-tailed example



10 arms, lomax loss
Opt. arm: CVaR = 2.55
Rest: CVaR = 3
$\alpha = 0.95$

Hard case



10 arms
Opt. arm: exponential, CVaR = 2.55
5 arms: lomax, CVaR = 3
4 arms: exponential, CVaR = 3
$\alpha = 0.95$

Concluding remarks

- Motivated environment oblivious, risk-aware MAB problem

- Pure exploration setting:
  - Two algorithm classes that outperform use of naïve empirical estimator
  - Prob. of error decays slower than exponentially in horizon length
  - Open: Fundamental lower bounds for the environment oblivious setting

- Open**: Environment oblivious regret minimization

Environment oblivious, risk-aware multi-armed banditry

Jayakrishnan Nair (IITB)

joint work with Anmol Kagrecha (IITB) & Krishna Jagannathan (IITM)