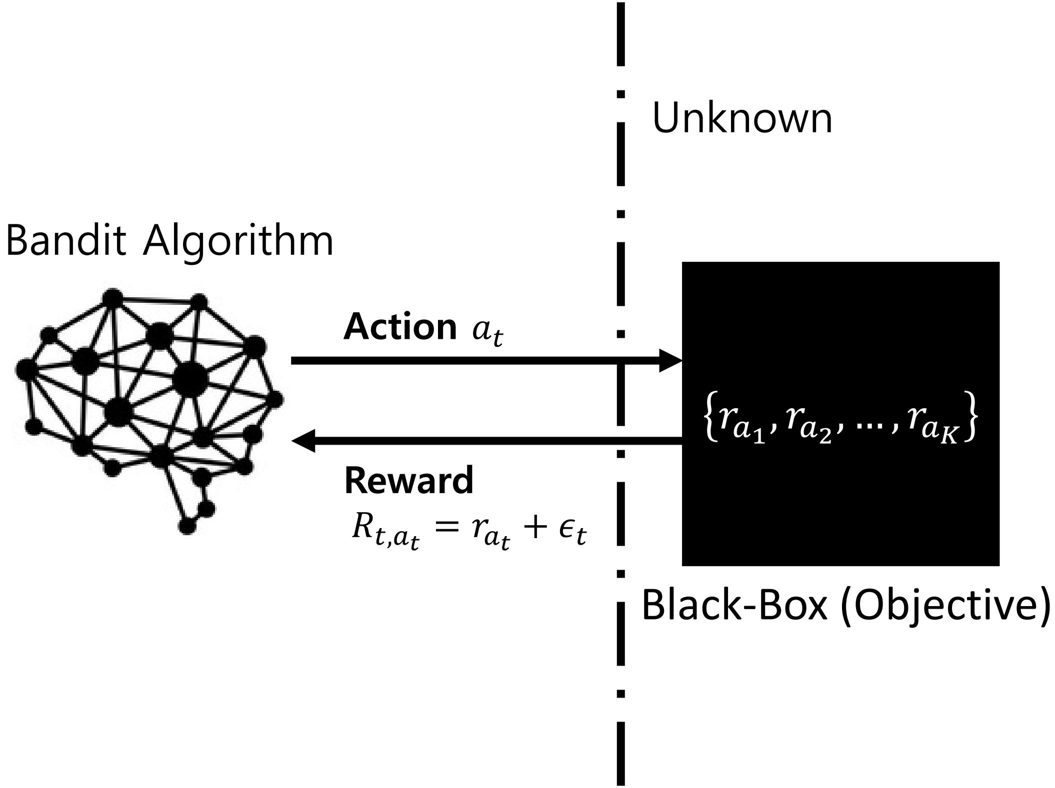
Optimal Algorithms for MABs with Heavy-Tailed Rewards

Presenter: Kyungjae Lee

(Stochastic) Multi-Armed Bandits

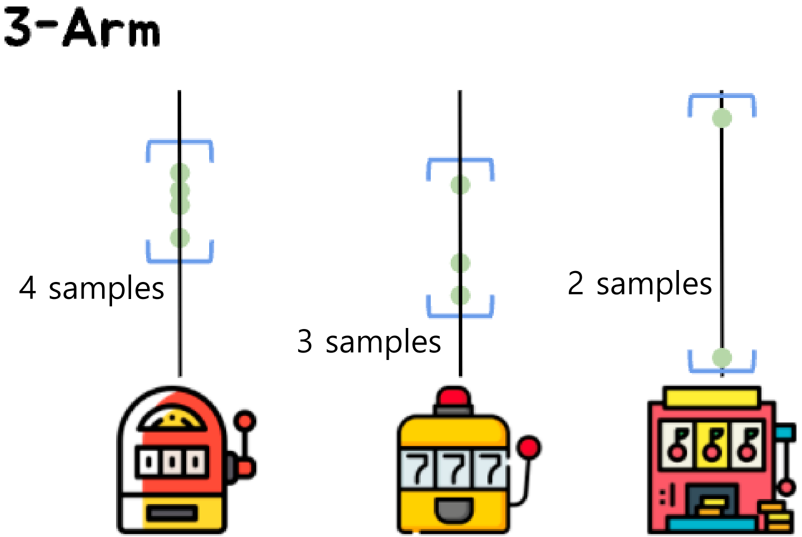
$$\max_{a \in \mathcal{A}} r_a$$

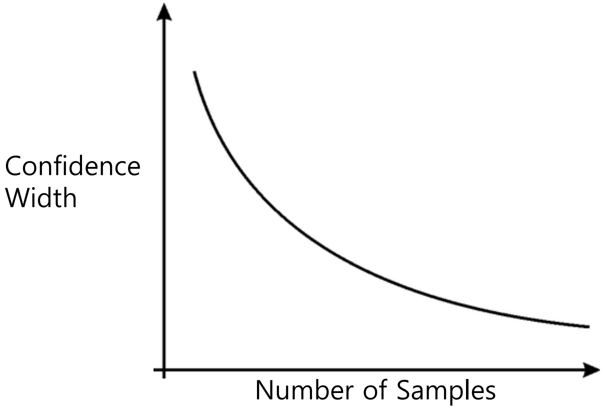
- Set of discrete actions
  - $\mathcal{A} = \{a_1, a_2, \dots, a_K\}$
- Mean rewards (**Unknown**)
  - $r_a \in [0,1]$
- I.I.D. Noise (**Unknown**)
  - $\epsilon_t \sim P_{\text{noise}}$



Structures of Upper Confidence Bounds

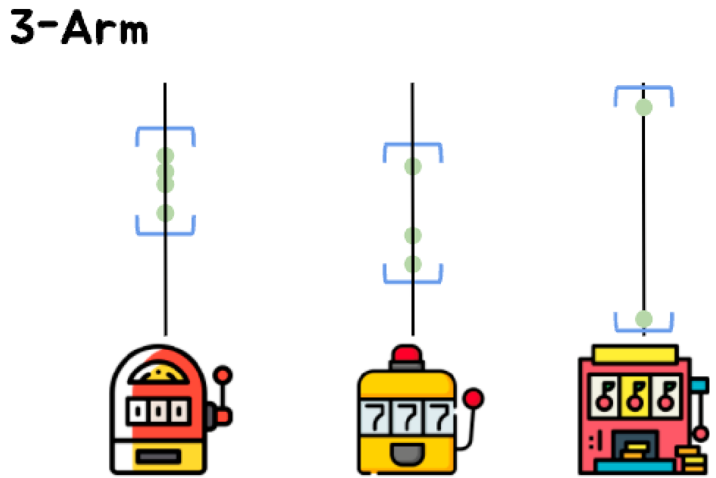
- Trade-off between *exploitation* and *exploration.*
  - *Exploitation*: choosing the best action
  - *Exploration*: gathering new information





Structures of Upper Confidence Bounds

- 1. Observations: $R_{t,a_t} = r_{a_t} + \epsilon_t$
- 2. Estimations: $\hat{r}_a = \frac{\sum_k^t R_{t,a_t} \mathbb{I}[a_t = a]}{n_a(t)}$
- 3. Confidence Bounds (or Estimation Error): $|r_a - \hat{r}_a| \leq C(n_a(t), \delta)$
- 4. Action Selection: $a_{t+1} := \text{argmax}_a \{\hat{r}_a + C(n_a(t), \delta)\}$

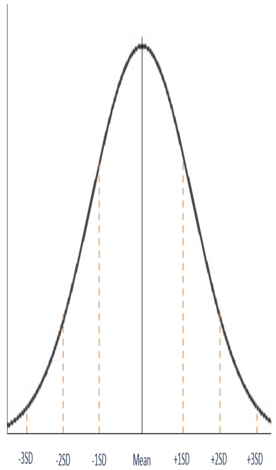


Basic Assumption on Noise Distribution

- Sub-Gaussian Noise (General assumption)
  - $\mathbb{E}\left[e^{\lambda \epsilon_t}\right] \leq e^{\lambda^2 \sigma^2}$
  - Error probability of sample mean estimator

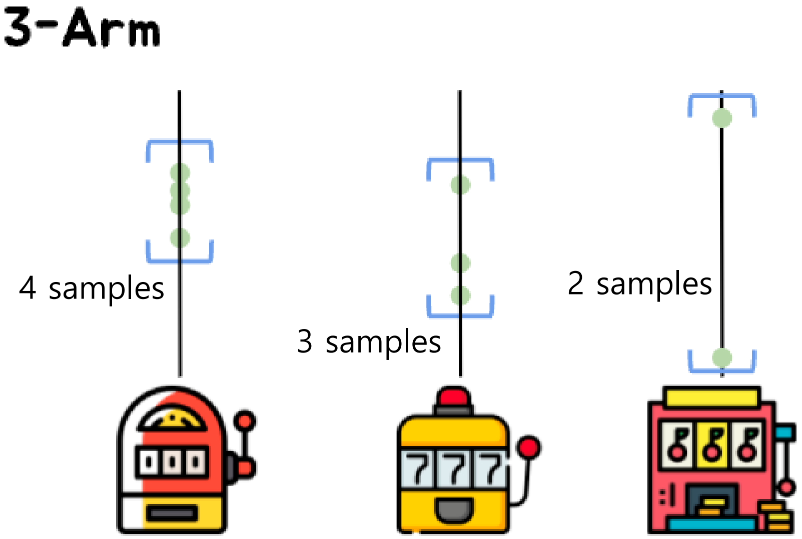
$$\mathbb{P}\left(\left|Y - \hat{Y}_n\right| > \epsilon\right) \leq \exp(-an\epsilon^2)$$

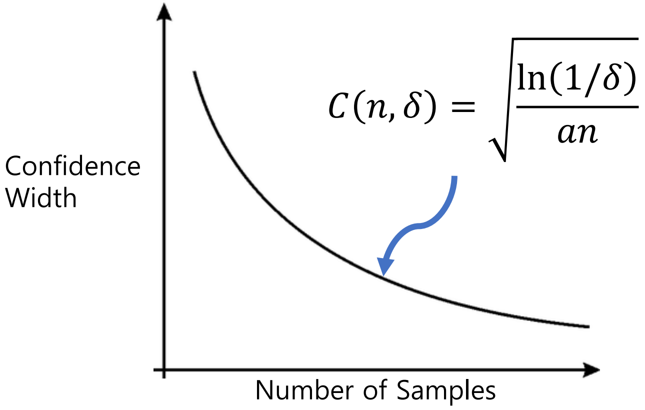
$$\mathbb{P}\left(\left|Y - \hat{Y}_n\right| > \sqrt{\frac{\ln(1/\delta)}{an}}\right) \leq \delta$$



Structures of Upper Confidence Bounds

- Trade-off between *exploitation* and *exploration.*
  - *Exploitation*: choosing the best action
  - *Exploration*: gathering new information



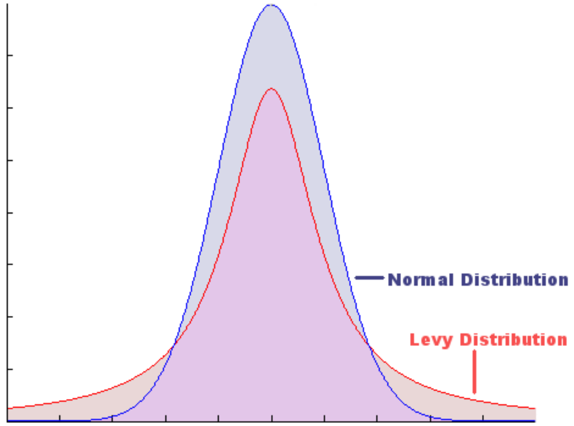


Bandits with Heavy-Tailed Rewards

- Sub-Gaussian Noise (General assumption)
  - $\mathbb{E}\left[e^{\lambda\epsilon_t}\right] \leq e^{\lambda^2\sigma^2}$
  - Error probability of sample mean estimator

$$\mathbb{P}\left(\left|Y - \hat{Y}_n\right| > \epsilon\right) \leq \exp(-an\epsilon^2)$$

- Heavy-tailed Noise (Bubeck et al., 2013)
  - $\mathbb{E}[|\boldsymbol{\epsilon_t}|^p] \leq \nu_p \quad (1 < p \leq 2)$
  - Exponential rate of sample mean estimator does not hold



Original Def.
$M_X(t) = \mathbb{E}[e^{tX}] = \infty$ for all $t > 0$

- Examples
  - Pareto, Log-normal, Weibull, Fréchet, etc.
  - Clinical trials, finance, delays in end-to-end network routing

Bubeck, Sébastien, Nicolo Cesa-Bianchi, and Gábor Lugosi. "Bandits with heavy tail." IEEE Transactions on Information Theory 59.11 (2013): 7711-7717.

Minimax Optimal Bandits for Heavy-Tailed Rewards

[1] Kyungjae Lee, Hongjun Yang, Sungbin Lim, and Songhwai Oh, "**Optimal Algorithms for Stochastic Multi-Armed Bandits with Heavy Tailed Rewards**," in Proc. of Neural Information Processing Systems (NeurIPS), Dec. 2020.
[2] Kyungjae Lee and Sungbin Lim, "**Minimax Optimal Bandits for Heavy Tail Rewards**," IEEE Transactions on Neural Networks and Learning Systems, 2022.

- Robust Estimator
- Drawbacks of Naïve Approaches
  - Robust UCB
  - Adaptively Perturbed Exploration (APE)
    - Unbounded Perturbation
- Optimal Strategies
  - **M**inimax Optimal **R**obust **U**pper **C**onfidence **B**ound (MR-UCB)
  - **M**inimax Optimal **R**obust **A**daptively **P**erturbed **E**xploration (MR-APE)
    - Bounded Perturbation

- Applications

Robust Estimator

- We propose a new robust estimator whose **error probability decays exponentially** and **does not require ν_p.** (inspired by Catoni 2012, Ceasa et al. 2017)

  - Influence function

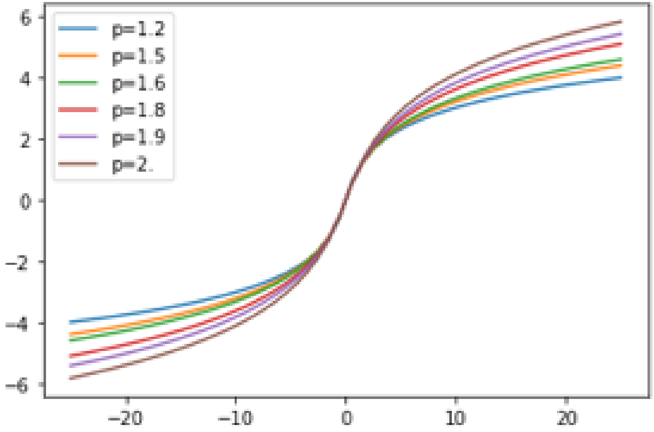
$$\psi_p(x) := \operatorname{sign}(x) \ln \left(b_p|x|^p + |x| + 1\right)$$

$Y + \epsilon \longrightarrow \psi(Y + \epsilon)$

Large noise → $\ln(Y + \epsilon)$

Small noise → $Y + \epsilon$

Influence function for different *p*



Cesa-Bianchi, Nicolò, et al. "Boltzmann exploration done right." Advances in neural information processing systems 30 (2017).
Catoni, Olivier. "Challenging the empirical mean and empirical variance: a deviation study." Annales de l'IHP Probabilités et statistiques. Vol. 48. No. 4. 2012.

Robust Estimator

- We propose a new robust estimator whose **error probability decays exponentially** and **does not require ν_p.** (inspired by Catoni 2012, Ceasa et al. 2017)

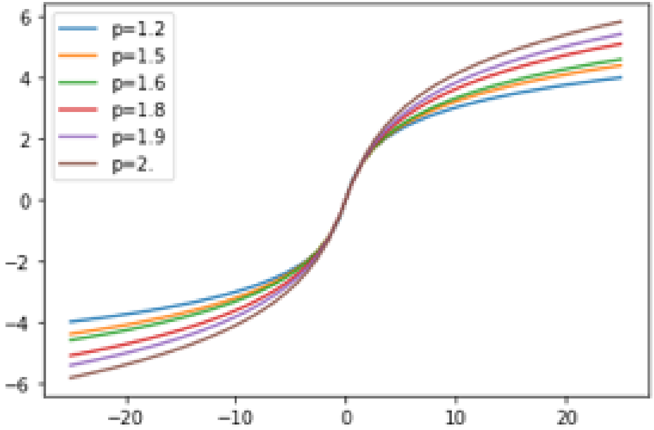
  - Influence function

$$\psi_p(x) := \operatorname{sign}(x) \ln\left(b_p|x|^p + |x| + 1\right)$$

  - p-Robust estimator

$$\hat{Y}_n := c/n^{1-1/p} \cdot \sum_{k=1}^{n} \psi_p\left(Y_k/(cn^{1/p})\right)$$

Influence function for different *p*



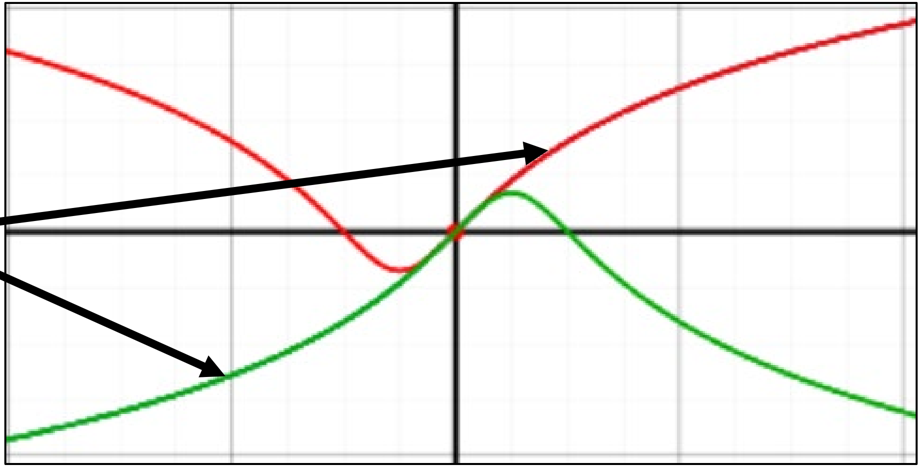
Cesa-Bianchi, Nicolò, et al. "Boltzmann exploration done right." Advances in neural information processing systems 30 (2017).
Catoni, Olivier. "Challenging the empirical mean and empirical variance: a deviation study." Annales de l'IHP Probabilités et statistiques. Vol. 48. No. 4. 2012.

Robust Estimator

Corollary 2. *Let* $b_p := \left[2\left(\frac{2-p}{p-1}\right)^{1-\frac{2}{p}} + \left(\frac{2-p}{p-1}\right)^{2-\frac{2}{p}}\right]^{-\frac{p}{2}}$. *For all* $x \in \mathbb{R}$*, the following inequality holds*

$$\ln\left(1 + x + b_p|x|^p\right) \geq -\ln\left(1 - x + b_p|x|^p\right).$$

$$\psi_p(x) := \text{sign}(x)\ln\left(b_p|x|^p + |x| + 1\right)$$



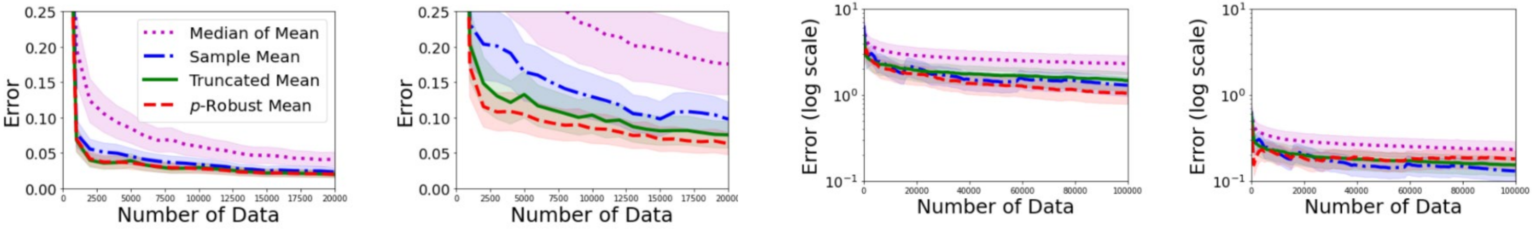
Robust Estimator

Theorem 2. *Let* $\{Y_k\}_{k=1}^{\infty}$ *be i.i.d. random variables sampled from a heavy-tailed distribution with a finite p-th moment,* $\nu_p := \mathbb{E}\,|Y_k|^p$*, for* $p \in (1, 2]$*. Let* $y := \mathbb{E}\,[Y_k]$ *and define an estimator as*

$$\hat{Y}_n := c/n^{1-1/p} \cdot \sum_{k=1}^{n} \psi_p\left(Y_k/(cn^{1/p})\right) \tag{4}$$

where $c > 0$ *is a constant. Then, for all* $\epsilon > 0$,

$$\mathbb{P}\left(\hat{Y}_n > y + \epsilon\right) \le \exp\left(-\frac{n^{\frac{p-1}{p}}\epsilon}{c} + \frac{b_p \nu_p}{c^p}\right),\ \mathbb{P}\left(y > \hat{Y}_n + \epsilon\right) \le \exp\left(-\frac{n^{\frac{p-1}{p}}\epsilon}{c} + \frac{b_p \nu_p}{c^p}\right). \tag{5}$$



(a) $p = 1.9, \lambda_\epsilon = 1.0$ (b) $p = 1.5, \lambda_\epsilon = 1.0$ (c) $p = 1.1, \lambda_\epsilon = 1.0$ (d) $p = 1.1, \lambda_\epsilon = 0.1$

Cumulative Regret

- Efficiency of Exploration
  - Minimize the cumulative regret over total rounds

$$\mathcal{R}_T \coloneqq \sum_{t}^{T} r_{a_\star} - \mathbb{E}\left[r_{a_t}\right]$$

  - The smaller the regret, the better the exploration performance
  - It is known that, for any algorithm,

$$\mathcal{R}_T \geq \Omega\left(K^{1-1/p}T^{1/p}\right) \quad (1 < p \leq 2)$$

For sub-Gaussian case,
$\mathcal{R}_T \geq \Omega\left(\sqrt{KT}\right)$

Bubeck, Sébastien, Nicolo Cesa-Bianchi, and Gábor Lugosi. "Bandits with heavy tail." IEEE Transactions on Information Theory 59.11 (2013): 7711-7717.

Robust Upper Confidence Bound

- Robust Upper Confidence Bound (Robust UCB, Bubeck et al., 2017)

$$a_t = \arg\max \ \hat{\boldsymbol{r}}_{\boldsymbol{t},\boldsymbol{a}} + \boldsymbol{v}_{\boldsymbol{p}} \left(\frac{\boldsymbol{\eta} \ln(\boldsymbol{t}^2)}{\boldsymbol{n}_{\boldsymbol{t},\boldsymbol{a}}} \right)^{1-\frac{1}{p}}$$

- Robust Estimators

Confidence bound of
truncated estimator, median of means

$$\hat{r} \leq r + v_p^{1/p} \left(\frac{\ln(1/\delta)}{n} \right)^{1-\frac{1}{p}}$$

The order stems from the error bound of the robust estimator

Robust Upper Confidence Bound

- Robust Upper Confidence Bound (Robust UCB, Bubeck et al., 2017)

$$a_t = \arg\max \ \hat{\boldsymbol{r}}_{t,a} + \boldsymbol{\nu}_p \left(\frac{\boldsymbol{\eta} \ln\left(\boldsymbol{t}^2\right)}{\boldsymbol{n}_{t,a}} \right)^{1-\frac{1}{p}}$$

- Regret Bounds

The instance dependent regret of the Robust UCB satisfies

$$\mathcal{R}_T \leq O\left(\left(\frac{\nu_p}{\Delta_a} \right)^{\frac{1}{p-1}} \ln(T) + \Delta_a \right)$$

Also, the minimax regret satisfies

$$\mathcal{R}_T \leq O\left(T^{1/p} \left(K \cdot \ln(T) \right)^{1-1/p} \right)$$

The order stems from the error bound of the robust estimator

Minimax Lower Bounds

RAI LAB
Robotics and Artificial Intelligence Lab

- Failure of Robust UCB (Lee et al. 2020)

Theorem 1 (in [8]): There exists a K-armed stochastic bandit problem for which the regret of robust UCB has the following lower bound, for $T > \max(10, [(\nu^{(1/(p-1))}/\eta(K-1))]^2)$:

$$\mathcal{R}_T \geq \Omega\left((K\ln(T))^{1-1/p}T^{1/p}\right). \quad (8)$$

$$\mathcal{R}_T \geq \Omega\left(K^{1-1/p}T^{1/p}\right)$$

A revision of the confidence bound is necessary

- Some observations in UCB and MOSS under the sub-Gaussian assumption

UCB1	$\min\left(\sqrt{nK\log n}\,,\ \sum_{i:\Delta_i>0}\frac{\log n}{\Delta_i}\right)$
MOSS	$\min\left(\sqrt{nK}\,,\ \sum_{i:\Delta_i>0}\frac{K\log(2+n\Delta_i^2/K)}{\Delta_i}\right)$
EXP3	$\sqrt{nK\log K}$
INF	$\sqrt{nK}$

Table 1: regret upper bounds (up to a numerical constant factor) for different policies in the multi-armed bandit problem.

Minimax optimality can be achieved by simply modifying the confidence bound.

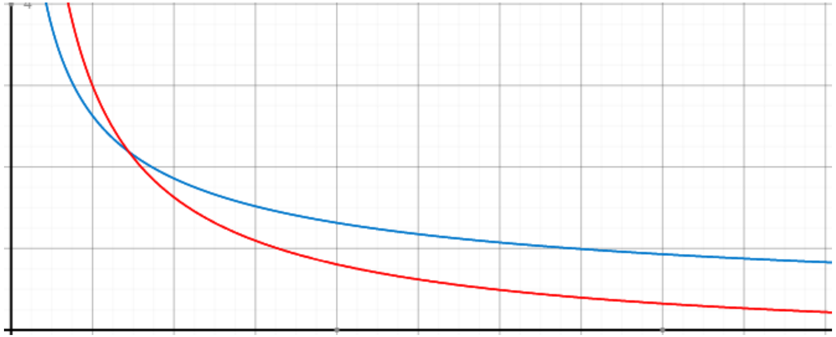
Optimal Strategies

RAI LAB
Robotics and Artificial Intelligence Lab

- **M**inimax Optimal **R**obust **U**pper **C**onfidence **B**ound (MR-UCB)
  - With our robust estimator

$$a_t = \arg\max \ \hat{\boldsymbol{r}}_{t,a} + \frac{c\ln_+\left(\frac{T}{K\cdot n_{t,a}}\right)}{n_{t,a}^{1-\frac{1}{p}}} \iff \nu_p\left(\frac{\eta\ln(t^2)}{n_{t,a}}\right)^{1-\frac{1}{p}}$$

Tendency of confidence bounds



MR-UCB

Robust-UCB

1. The order of logarithm is changed
2. The logarithm term is modified
3. ν_p can be unknown

Adaptively Perturbed Exploration

RAI LAB
Robotics and Artificial Intelligence Lab

- Adaptively Perturbed Exploration with a p-Robust Estimator (APE)

$$a_t = \arg\max \ \hat{r}_{t,a} + \beta_{t,a} G_{t,a}$$

Robust Estimation

Adaptive Coefficient

Random Perturbation

- $\beta_{t,a} := \frac{c}{n_{t,a}^{1-\frac{1}{p}}}$: adaptive coefficient that gradually decays
- $G_{t,a}$: random perturbation sampled from CDF $F(g)$

The order stems from the error bound of the robust estimator

Adaptively Perturbed Exploration

- Regret Upper Bound for Perturbation *G* with CDF *F(g)*

Theorem 3. *Assume that the p-th moment of rewards is bounded by a constant $\nu_p < \infty$, $\hat{r}_{t,a}$ is a p-robust estimator of (4) and $F(x)$ satisfies Assumption 2. Then, $\mathbb{E}[\mathcal{R}_T]$ of APE² is bounded as*

$$O\left(\sum_{a \neq a^\star} \frac{C_{p,\nu_p,F}}{\Delta_a^{\frac{1}{p-1}}} + \frac{(6c)^{\frac{p}{p-1}}}{\Delta_a^{\frac{1}{p-1}}} \left[-F^{-1}\left(\frac{c^{\frac{p}{p-1}}}{T\Delta_a^{\frac{p}{p-1}}} \right) \right]_+^{\frac{p}{p-1}} + \frac{(3c)^{\frac{p}{p-1}}}{\Delta_a^{\frac{1}{p-1}}} \left[F^{-1}\left(1 - \frac{c^{\frac{p}{p-1}}}{T\Delta_a^{\frac{p}{p-1}}} \right) \right]_+^{\frac{p}{p-1}} + \Delta_a \right)$$

where $[x]_+ := \max(x, 0)$, $C_{p,\nu_p,F} > 0$ is a constant independent of T.

  - *the upper bound can be calculated using a plug-and-play manner*

Adaptively Perturbed Exploration

RAI LAB
Robotics and Artificial Intelligence Lab

- Regret Upper Bound for Perturbation *G* with CDF *F(g)*

Theorem 3. *Assume that the p-th moment of rewards is bounded by a constant $\nu_p < \infty$, $\hat{r}_{t,a}$ is a p-robust estimator of (4) and $F(x)$ satisfies Assumption 2. Then, $\mathbb{E}[\mathcal{R}_T]$ of APE[2] is bounded as*

$$O\left(\sum_{a \neq a^\star} \frac{C_{p,\nu_p,F}}{\Delta_a^{\frac{1}{p-1}}} + \frac{(6c)^{\frac{p}{p-1}}}{\Delta_a^{\frac{1}{p-1}}}\left[-F^{-1}\left(\frac{c^{\frac{p}{p-1}}}{T\Delta_a^{\frac{p}{p-1}}}\right)\right]_+^{\frac{p}{p-1}} + \frac{(3c)^{\frac{p}{p-1}}}{\Delta_a^{\frac{1}{p-1}}}\left[F^{-1}\left(1-\frac{c^{\frac{p}{p-1}}}{T\Delta_a^{\frac{p}{p-1}}}\right)\right]_+^{\frac{p}{p-1}} + \Delta_a\right)$$

where $[x]_+ := \max(x, 0)$, $C_{p,\nu_p,F} > 0$ is a constant independent of T.

  - *the upper bound can be calculated using a plug-and-play manner*

The estimator is poorly concentrated

Good estimation and small perturbation, yet the sub-optimal arm is selected

Good estimation is given, but the sub-optimal arm is selected due to the large perturbation

Adaptively Perturbed Exploration

- Regret Analysis for various perturbations

Dist. on G	Prob. Dep. Bnd. $O(\cdot)$	Prob. Indep. Bnd. $O(\cdot)$	Low. Bnd. $\Omega(\cdot)$	Opt. Params.	Opt. Bnd. $\Theta(\cdot)$
Weibull	$\sum_{a\neq a^\star} A_{c,\lambda,a} \left(\ln\left(B_{c,a}T\right)\right)^{\frac{p}{k(p-1)}}$	$C_{K,T} \ln\left(K\right)^{\frac{1}{k}}$	$C_{K,T} \ln\left(K\right)$	$k=1, \lambda \geq 1$	
Gamma	$\sum_{a\neq a^\star} A_{c,\lambda,a} \alpha^{p/(p-1)} \ln\left(B_{c,a}T\right)^{p/(p-1)}$	$C_{K,T} \frac{\ln\left(\alpha K^{1+p/(p-1)}\right)^{p/(p-1)}}{\ln(K)^{\frac{1}{p-1}}}$	$C_{K,T} \ln\left(K\right)$	$\alpha=1, \lambda \geq 1$	$K^{1-1/p}T^{1/p} \ln\left(K\right)$
GEV	$\sum_{a\neq a^\star} A_{c,\lambda,a} \ln_\zeta\left(B_{c,a}T\right)^{p/(p-1)}$	$C_{K,T} \frac{\ln_\zeta\left(K^{\frac{2p-1}{p-1}}\right)^{p/(p-1)}}{\ln_\zeta(K)^{\frac{1}{p-1}}}$	$C_{K,T} \ln_\zeta\left(K\right)$	$\zeta=0, \lambda \geq 1$	
Pareto	$\sum_{a\neq a^\star} A_{c,\lambda,a} \left[B_{c,a}T\right]^{\frac{p}{\alpha(p-1)}}$	$C_{K,T} \alpha^{1+\frac{p^2}{\alpha(p-1)^2}} K^{\frac{1}{\alpha(p-1)}}$	$C_{K,T} \alpha K^{\frac{1}{\alpha}}$	$\alpha=\lambda=\ln(K)$	
Fréchet	$\sum_{a\neq a^\star} A_{c,\lambda,a} \left[B_{c,a}T\right]^{\frac{p}{\alpha(p-1)}}$	$C_{K,T} \alpha^{1+\frac{p^2}{\alpha(p-1)^2}} K^{\frac{1}{\alpha(p-1)}}$	$C_{K,T} \alpha K^{\frac{1}{\alpha}}$	$\alpha=\lambda=\ln(K)$	

- For any perturbation, we achieve an optimal regret bound with respect to T

Minimax Lower Bounds

- Failure of APE

Theorem 4. *For* $0 < c < \frac{K-1}{K-1+2^{p/(p-1)}}$ *and* $T \geq \frac{c^{1/(p-1)}(K-1)}{2^{p/(p-1)}} \left|F^{-1}\left(1 - \frac{1}{K}\right)\right|^{p/(p-1)}$, *there exists a* K*-armed stochastic bandit problem where the regret of* APE^2 *is lower bounded by* $\mathbb{E}[\mathcal{R}_T] \geq \Omega\left(K^{1-1/p}T^{1/p}F^{-1}\left(1 - 1/K\right)\right)$.

$\mathcal{R}_T \geq \Omega\left(K^{1-1/p}T^{1/p}\right)$ } The support of G must be bounded.

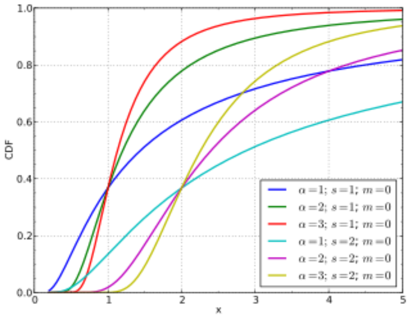
Minimax Lower Bounds

- Failure of APE

Theorem 4. *For* $0 < c < \frac{K-1}{K-1+2^{p/(p-1)}}$ *and* $T \geq \frac{c^{1/(p-1)}(K-1)}{2^{p/(p-1)}} \left|F^{-1}\left(1-\frac{1}{K}\right)\right|^{p/(p-1)}$, *there exists a* K*-armed stochastic bandit problem where the regret of APE*^{2} *is lower bounded by* $\mathbb{E}[\mathcal{R}_T] \geq \Omega\left(K^{1-1/p}T^{1/p}F^{-1}\left(1-1/K\right)\right)$.

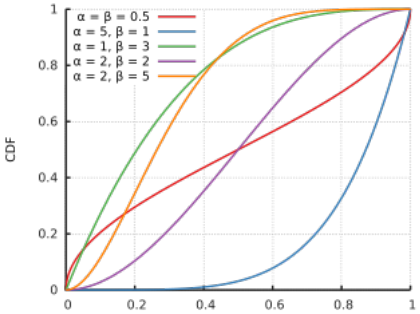
$\mathcal{R}_T \geq \Omega\left(K^{1-1/p}T^{1/p}\right)$ } The support of G must be bounded.

$F^{-1}(1-1/K) \in [0, \infty]$



CDF of Frechet

$F^{-1}(1-1/K) \in [A, B]$



Beta of Frechet

Optimal Strategies

- **M**inimax Optimal **R**obust **U**pper **C**onfidence **B**ound (MR-UCB)

$$a_t = \arg\max \ \hat{\boldsymbol{r}}_{t,a} + \frac{\mathbf{c}\ln_+\left(\frac{\boldsymbol{T}}{\boldsymbol{K}\cdot\boldsymbol{n}_{t,a}}\right)}{\boldsymbol{n}_{t,a}^{1-\frac{1}{p}}}$$

- **M**inimax Optimal **R**obust **A**daptively **P**erturbed **E**xploration (MR-APE)

$$a_t = \arg\max \ \hat{\boldsymbol{r}}_{t,a} + (\mathbf{1}+\boldsymbol{\epsilon})\boldsymbol{\beta}_{t,a}\boldsymbol{G}_{t,a}$$

  - $\boldsymbol{G}_{t,a}$: bounded random perturbation

Cumulative Regret Bounds

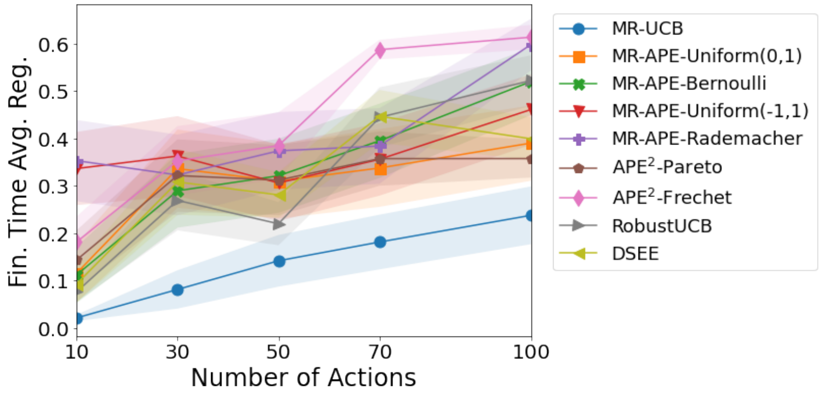
- Regret Analysis for Minimax Optimal Algorithms (Lee and Lim, 2022)
  - Modified Confidence Bounds

Algorithm		Gap-Dependent Bound $O(\cdot)$	Gap-Independent Bound $\Theta(\cdot)$	Prior Info.
Robust UCB [7]		$\sum_{a\neq a_\star} \ln(T)/\Delta_a^{1/(p-1)}$	$(K\ln(T))^{1-\frac{1}{p}}T^{\frac{1}{p}}$	p and ν_p
Robust MOSS [9]		$\sum_{a\neq a_\star} \ln\left(T\Delta_a^{p/(p-1)}/K\right)/\Delta_a^{1/(p-1)}$	$K^{1-\frac{1}{p}}T^{\frac{1}{p}}$	
MR-UCB [This work]		$\sum_{a\neq a_\star} \ln\left(T\Delta_a^{p/(p-1)}/K\right)^{p/(p-1)}/\Delta_a^{1/(p-1)}$	$K^{1-\frac{1}{p}}T^{\frac{1}{p}}$	p
APE^2 (Unbounded) [8]	TYPE I	$\sum_{a\neq a_\star} \ln\left(T\Delta_a^{p/(p-1)}\right)^{p/(p-1)}/\Delta_a^{1/(p-1)}$	$K^{1-\frac{1}{p}}T^{\frac{1}{p}}\ln(K)$	
	TYPE II	$\sum_{a\neq a_\star} \ln(K)^{\frac{p}{p-1}}\left(T\Delta_a^{p/(p-1)}\right)^{\frac{p}{\ln(K)(p-1)}}/\Delta_a^{1/(p-1)}$		
MR-APE^2 (Bounded) [This work]		$\sum_{a\neq a_\star} \ln\left(T\Delta_a^{p/(p-1)}/K\right)^{p/(p-1)}/\Delta_a^{1/(p-1)}$	$K^{1-\frac{1}{p}}T^{\frac{1}{p}}$	

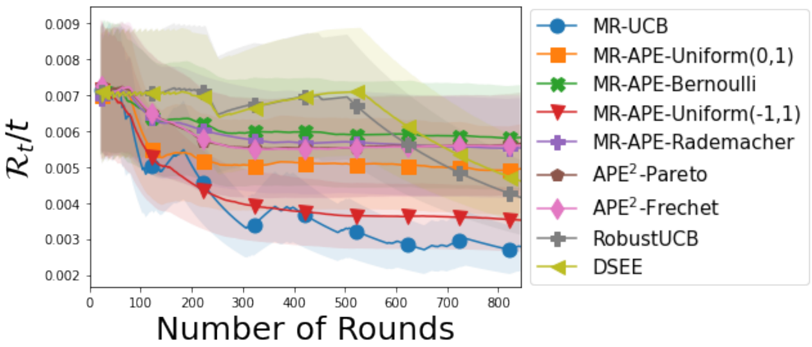
Simulations

- Effect of the number of action (K)
  - Gap: 0.7
  - Noise: Pareto distribution

Algorithm		Gap-Independent Bound $\Theta(\cdot)$
Robust UCB [7]		$(K\ln(T))^{1-\frac{1}{p}}T^{\frac{1}{p}}$
Robust MOSS [9]		$K^{1-\frac{1}{p}}T^{\frac{1}{p}}$
MR-UCB [This work]		$K^{1-\frac{1}{p}}T^{\frac{1}{p}}$
APE^2 (Unbounded) [8]	TYPE I	$K^{1-\frac{1}{p}}T^{\frac{1}{p}}\ln(K)$
	TYPE II	
MR-APE^2 (Bounded) [This work]		$K^{1-\frac{1}{p}}T^{\frac{1}{p}}$



- Performance on cryptocurrency dataset



Future Extension

- Linear Contextual Bandits with Heavy-Tailed Noise
  - Robust linear estimator! and its error bound, but nothing special…
- Bayesian Optimization with Heavy-Tailed Noise
  - Robust kernel ridge estimator! and its error bound, some improvement can be achieved!
  - Submitted to ECAI 2024
- Beyond the moment assumption and mean estimation
  - Mode estimation (Pacchiano et al. 2021), Quantile estimation (Zhang and Cheng, 2021)
  - Risk estimation (Vincent et al. 2022, Saux and Maillard, 2023)

Pacchiano, Aldo, Heinrich Jiang, and Michael I. Jordan. "Robustness Guarantees for Mode Estimation with an Application to Bandits." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 10. 2021.
Zhang, Mengyan, and Cheng Soon Ong. "Quantile bandits for best arms identification." International Conference on Machine Learning. PMLR, 2021.
Vincent Y. F. Tan, Prashanth L. A., Krishna P. Jagannathan: A Survey of Risk-Aware Multi-Armed Bandits. IJCAI 2022.
Saux, Patrick, and Odalric Maillard. "Risk-aware linear bandits with convex loss." International Conference on Artificial Intelligence and Statistics. PMLR, 2023.

Thank you!