清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

Adaptive Best-of-Both-Worlds Algorithm for Heavy-Tailed Multi-Armed Bandits

Longbo Huang
Tsinghua University

2022 MSRA Data-Driven Optimization Workshop

The Multi-Armed Bandit Model

The MAB framework has many applications

- Online advertising
- Wireless communication
- Clinical trial
- Recommender system
- AI technology
- Finance
- …

Extensively studied

[Auer et al. 2002a][Auer et al 2002b][Agrawal&Goyal 2012][Bubeck&Cesa-Bianchi 2012][Tao et al. 2018][Kuroki et al. 2020][Du et al. 2020] [Wang&Chen 2022][Chen&Zhao&Li 2022] …

Multi-Armed Bandits

The MAB problem

- T time steps, K actions ("arms")
- $\{l_{t,a}\}$: $T \times K$ loss matrix
- Each time we choose A_t, suffer & observe a loss l_{t,A_t}
- Minimize "pseudo-regret"

$$\max_{a \in [K]} \mathbb{E}\left[\sum_{t=1}^{T} l_{t,A_t} - \sum_{t=1}^{T} l_{t,a}\right]$$

Heavy-Tailed Multi-Armed Bandits

The Heavy-Tailed MAB problem

- T time steps, K actions ("arms")
- $\{l_{t,a}\}$: $T \times K$ loss matrix
- Adversary picks (α, σ)-heavy-tailed distributions $\nu_{t,1}, \dots, \nu_{t,K}$ with

$$\mathbb{E}_{X\sim\nu}[|X|^\alpha] \leq \sigma^\alpha, 1 < \alpha \leq 2$$

- Each time we choose A_t, suffer & observe a loss $l_{t,A_t} \sim \nu_{t,A_t}$
- Minimize "pseudo-regret"

$$\max_{a\in[K]} \mathbb{E}\left[\sum_{t=1}^{T} l_{t,A_t} - \sum_{t=1}^{T} l_{t,a}\right]$$

Heavy-Tailed Multi-Armed Bandits

The Heavy-Tailed MAB problem

- T time steps, K actions ("arms")
- $\{l_{t,a}\}$: $T \times K$ loss matrix
- Adversary picks (α, σ)-heavy-tailed distributions $\nu_{t,1}, \dots, \nu_{t,K}$ with

$$\mathbb{E}_{X \sim \nu}[|X|^\alpha] \leq \sigma^\alpha, 1 < \alpha \leq 2$$

Natural generalization in both directions

- Stochastic heavy-tailed MAB: $\nu_{1,a} = \nu_{2,a} = \cdots \nu_{T,a}$
- Classical Adversarial MAB: $\nu_{t,a}$ is a Dirac-measure at $l_{t,a} \in [0,1]$

Heavy-Tailed Multi-Armed Bandits

The Heavy-Tailed MAB problem

- T time steps, K actions ("arms")
- $\{l_{t,a}\}$: $T \times K$ loss matrix
- Adversary picks (α, σ)-heavy-tailed distributions $\nu_{t,1}, \dots, \nu_{t,K}$ with

$$\mathbb{E}_{X \sim \nu}[|X|^\alpha] \leq \sigma^\alpha, 1 < \alpha \leq 2$$

Challenges

- Potentially unbounded 2nd moment (estimation and concentration)
- Unknown σ and α values (learning)

Representative Results on Heavy-Tailed MAB

Algorithm	Loss Type	Prior Knowledge	Total Regret
Lower-bounds (Bubeck et al., 2013)	Stochastic[a]	α, σ	$\Omega\left(\sigma^{\frac{\alpha}{\alpha-1}} \sum_{i\neq i^*} \Delta_i^{-\frac{1}{\alpha-1}} \log T\right)$
			$\Omega\left(\sigma K^{1-1/\alpha} T^{1/\alpha}\right)$
`RobustUCB` (Bubeck et al., 2013)	Stochastic	α, σ	$\mathcal{O}\left(\sum_{i\neq i^*} (\frac{\sigma^\alpha}{\Delta_i})^{\frac{1}{\alpha-1}} \log T\right)$ **(optimal)**
			$\mathcal{O}\left(\sigma (K \log T)^{1-1/\alpha} T^{1/\alpha}\right)$ **(sub-optimal for** $\log T$ **factors)**
Lee et al. (2020)	Stochastic	α; require $\mu_i \in [0,1]$	$\mathcal{O}\left(K^{1-1/\alpha} T^{1/\alpha} \log K\right)$[b] **(sub-optimal for** $\log K$ **factors)**
`1/2-Tsallis-INF` (Zimmert & Seldin, 2019)	SCA-unique[c]	$[0,1]$-bounded losses	$\mathcal{O}\left(\sum_{i\neq i^*} \frac{1}{\Delta_i} \log T\right)$ **(optimal for** $\alpha = 2, \sigma = 1$ **case)**
	Adversarial		$\mathcal{O}\left(\sqrt{KT}\right)$ **(optimal for** $\alpha = 2, \sigma = 1$ **case)**
`HTINF` **(ours)**	SCA-unique	α, σ	$\mathcal{O}\left(\sum_{i\neq i^*} (\frac{\sigma^\alpha}{\Delta_i})^{\frac{1}{\alpha-1}} \log T\right)$ **(optimal)**
	Adversarial		$\mathcal{O}\left(\sigma K^{1-1/\alpha} T^{1/\alpha}\right)$ **(optimal)**
Optimistic `HTINF` **(ours)**	SCA-unique	None	$\mathcal{O}\left(\sum_{i\neq i^*} \left(\frac{\sigma^{2\alpha}}{\Delta_i^{3-\alpha}}\right)^{\frac{1}{\alpha-1}} \log T\right)$
	Adversarial		$\mathcal{O}(\sigma^\alpha K^{\frac{\alpha-1}{2}} T^{\frac{3-\alpha}{2}})$
`AdaTINF` **(ours)**	Adversarial	None[d]	$\mathcal{O}\left(\sigma K^{1-1/\alpha} T^{1/\alpha}\right)$ **(optimal)**

Need to know σ, α before-hand

Our Contributions

Three novel algorithms

- Heavy-Tail Tsallis-INF (HTINF) – known σ, α
  - First to achieve *best-of-both-worlds* for heavy-tailed MAB
  - Applicable to unknown σ, α case (OptTINF): $O(\log T)$ for stochastic and $O(K^{\frac{\alpha-1}{2}} T^{\frac{3-\alpha}{2}})$ for adversarial
- Adaptive Tsallis-INF (AdaTINF) – zero knowledge
  - Optimal $O\left(\sigma K^{1-\frac{1}{\alpha}} T^{\frac{1}{\alpha}}\right)$ regret for adversarial

Heavy-Tail Tsallis-INF (HTINF) – Known σ, α

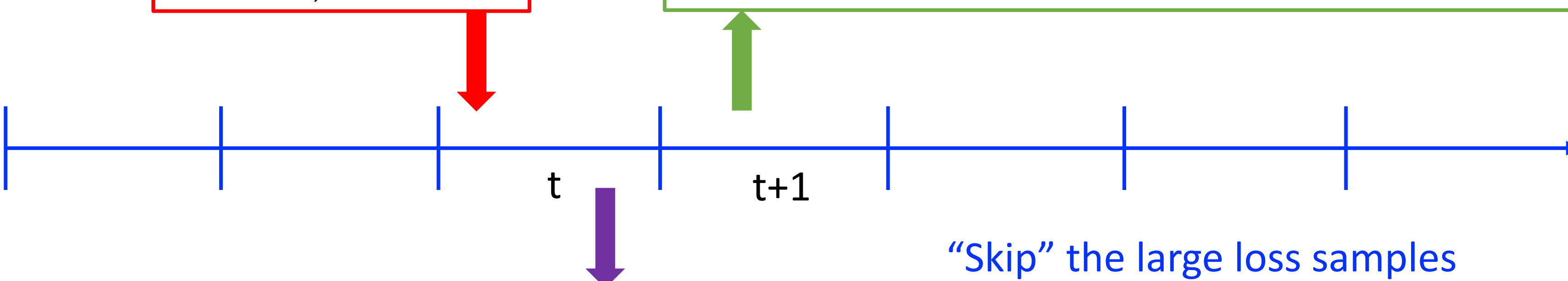
- Based on Follow-the-Regularized-Leader (FTRL)
- A novel skipping idea to "clip" large samples

3. Form x_{t+1} with **FTRL**:

$$x_{t+1} = \arg\min_{x \in \Delta_{[K]}} \left(\eta_{t+1} \left\langle \sum_{s=1}^{t} \hat{l}_s , x \right\rangle + \Psi(x) \right)$$

where $\Psi(x) \coloneqq -\alpha \sum_{i=1}^{K} x_i^{1/\alpha}$ (1/α-Tsallis entropy)

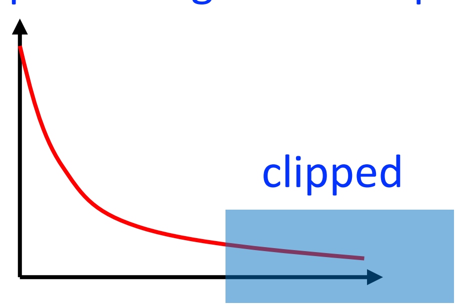
1. Pull arm a with prob. $x_{t,a}$ at t



2. Receive l_{t,A_t}, form $\hat{l}_{t,a}$ with **importance sampling** and **skipping**:

$$\hat{l}_{t,a} \coloneqq \mathbb{I}\left[A_t = a, \left|l_{t,A_t}\right| \le r_t\right] l_{t,A_t} / x_{t,A_t}$$

"Skip" the large loss samples



Heavy-Tail **T**sallis-**INF** (HTINF) – Known σ, α

Algorithm 1 `Heavy-Tail Tsallis-INF` (HTINF)

Input: Number of arms K, heavy-tail parameters α and σ

Output: Sequence of actions $i_1, i_2, \cdots, i_T \in [K]$

for $t = 1, 2, \cdots$ **do**

Calculate policy with learning rate $\eta_t^{-1} = \sigma t^{1/\alpha}$;

Pick the regularizer $\boxed{\Psi(x) = -\alpha \sum_{i=1}^{K} x_i^{1/\alpha}}$: 1/$\alpha$-Tsallis entropy

$$\boxed{x_t \leftarrow \operatorname*{argmin}_{x \in \triangle_{[K]}} \left(\eta_t \sum_{s=1}^{t-1} \langle \hat{\ell}_s, x \rangle + \Psi(x) \right)}$$

Follow-the-Regularized-Leader

Sample new action $i_t \sim x_t$.

Calculate the skipping threshold $\boxed{r_t \leftarrow \Theta_\alpha \eta_t^{-1} x_{t,i_t}^{1/\alpha}}$ Skipping threshold (avoid overly large loss in estimation)

where $\Theta_\alpha = \min\{1 - 2^{-\frac{\alpha-1}{2\alpha-1}}, (2 - \frac{2}{\alpha})^{\frac{1}{2-\alpha}}\}$.

Play according to i_t and observe loss feedback ℓ_{t,i_t}.

if $|\ell_{t,i_t}| > r_t$ **then**

$\hat{\ell}_t \leftarrow \mathbf{0}$.

else

Construct weighted importance sampling loss estimator $\boxed{\hat{\ell}_{t,i} \leftarrow \frac{\ell_{t,i}}{x_{t,i}} \mathbb{1}[i = i_t], \forall i \in [K].}$ Importance Sampling (biased due to skipping)

end if

end for

Heavy-Tail Tsallis-INF (HTINF) – Known σ, α

Theorem (Informal) HTINF achieves

- Adversarial environment

$$R_T \leq O\left(K^{1-\frac{1}{\alpha}} T^{\frac{1}{\alpha}} \log K\right)$$

- Stochastic environment

$$R_T \leq O\left(\sigma^{\frac{\alpha}{\alpha-1}} \sum_{a \neq a^*} \Delta_i^{-\frac{1}{\alpha-1}} \log T\right)$$

Remark

- **Best-of-both-worlds:** Both cases are optimal without knowing which environment beforehand

Adaptivity to unknown σ, α

Theorem (Informal) When σ, α are unknown, running HTINF with $\sigma = 1, \alpha = 2$ (OptTINF) achieves

- Adversarial environment

$$R_T \leq O\left(\sigma^{\alpha} K^{\frac{\alpha-1}{2}} T^{\frac{3-\alpha}{2}} + \sqrt{KT}\right)$$

- Stochastic environment

$$R_T \leq O\left(\sigma^{\frac{2\alpha}{\alpha-1}} \sum_{a \neq a^*} \Delta_i^{-\frac{3-\alpha}{\alpha-1}} \log T\right)$$

Remarks

- Still $O(\log T)$ regret for stochastic case
- o(T) regret for adversarial case

HTINF Regret Analysis

Regret decomposition of HTINF

$$\mathcal{R}_T(y) \triangleq \sum_{t=1}^{T} \mathbb{E}\left[\langle x_t - y, \mu_t \rangle\right] \quad (y \in \triangle_{[K]})$$

$$= \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \langle x_t - y, \mu_t - \mu_t' \rangle\right]}_{\text{Skipping gap}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \langle x_t - y, \hat{\ell}_t \rangle\right]}_{\text{FTRL Error}}$$

where $\mu_{t,i}' \triangleq \mathbb{E}[\ell_{t,i} \mathbb{1}[|\ell_{t,i}| \le r_t] \mid \mathcal{F}_{t-1}, i_t = i]$ is the clipped expectation

Skipping threshold impacts the regret

- A larger r_t leads to a smaller skipping gap but a larger FTRL error
- Optimal tradeoff achieved at $r_t = \Theta(\eta_t^{-1} x_{t,i_t}^{1/\alpha})$

HTINF Regret Analysis

Regret decomposition of HTINF

$$\mathcal{R}_T(y) \triangleq \sum_{t=1}^{T} \mathbb{E}\left[\langle x_t - y, \mu_t \rangle\right] \quad (y \in \triangle_{[K]})$$

$$= \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \langle x_t - y, \mu_t - \mu_t' \rangle\right]}_{\text{Skipping gap}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \langle x_t - y, \hat{\ell}_t \rangle\right]}_{\text{FTRL Error}}$$

where $\mu_{t,i}' \triangleq \mathbb{E}[\ell_{t,i} \mathbb{1}[|\ell_{t,i}| \le r_t] \mid \mathcal{F}_{t-1}, i_t = i]$ is the clipped expectation

Analysis idea

- "Self-bounding property" similar to (Zimmert & Seldin, 2019)
- *First* result for $1/\alpha$-Tsallis ($\alpha < 2$) regularized FTRL

HTINF Regret Analysis

Regret decomposition of HTINF

$$\mathcal{R}_T(y) = \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\langle x_t - y, \mu_t - \mu_t'\rangle\right]}_{\text{Skipping gap; By choice of } r_t \;\le 10\sigma(T+1)^{1/\alpha}K^{1-1/\alpha}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\langle x_t - y, \hat{\ell}_t\rangle\right]}_{\text{FTRL Error}}$$

$$= \underbrace{\sum_{t=1}^{T}\eta_t^{-1}D_\Psi(x_t, z_t)}_{\text{Part B; By Bregman divergence} \;\le 8\sigma t^{1/\alpha-1}K^{1-1/\alpha}} + \underbrace{\sum_{t=1}^{T}(\eta_t^{-1} - \eta_{t-1}^{-1})\left(\Psi(y) - \Psi(x_t)\right)}_{\text{Part A; By Tsallis Entropy} \;\le 4\sigma(T+1)^{1/\alpha}K^{1-1/\alpha}}$$

Adversarial case

Instance-independent

$R_T \le O\left(\sigma^\alpha K^{\frac{\alpha-1}{2}} T^{\frac{3-\alpha}{2}}\right)$

HTINF Regret Analysis

Regret decomposition of HTINF

$$\mathcal{R}_T(y) = \underbrace{\overbrace{\mathbb{E}\left[\sum_{t=1}^{T}\langle x_t - y, \mu_t - \mu_t'\rangle\right]}^{\text{Skipping gap}}}_{\text{By choice of } r_t} + \underbrace{\overbrace{\mathbb{E}\left[\sum_{t=1}^{T}\langle x_t - y, \hat{\ell}_t\rangle\right]}^{\text{FTRL Error}}}$$

$$\leq 5\sigma \sum_{t=1}^{T}\sum_{i \neq i^*} t^{1/\alpha - 1} x_{t,i}^{1/\alpha}$$

$$=$$

$$\overbrace{\underbrace{\sum_{t=1}^{T}\eta_t^{-1} D_\Psi(x_t, z_t)}_{\text{By Bregman divergence}}}^{\text{Part B}} + \overbrace{\underbrace{\sum_{t=1}^{T}(\eta_t^{-1} - \eta_{t-1}^{-1})\left(\Psi(y) - \Psi(x_t)\right)}_{\text{By Tsallis Entropy}}}^{\text{Part A}}$$

$$\leq 8\sigma t^{1/\alpha - 1}\sum_{i \neq i^*} x_{t,i}^{1/\alpha} \qquad \leq \sum_{t=1}^{T} 2\sigma t^{1/\alpha - 1}\sum_{i \neq i^*} x_{t,i}^{1/\alpha}$$

Stochastic case

Instance-dependent

$R_T \leq O(\log T)$

Adaptive Tsallis-INF (AdaTINF) – Unknown σ, α

Now consider not having σ, α

Skipping gap

FTRL Error = Part A + Part B

$$\mathcal{R}_T(y) = \mathbb{E}\left[\sum_{t=1}^{T} \langle x_t - y, \mu_t - \mu'_t \rangle\right] +$$

Decreasing in skipping threshold

Increasing in skipping threshold

Part B

Part A

$$\sum_{t=1}^{T} \eta_t^{-1} D_\Psi(x_t, z_t) + \sum_{t=1}^{T} (\eta_t^{-1} - \eta_{t-1}^{-1}) \left(\Psi(y) - \Psi(x_t)\right)$$

AdaTINF Idea: adjust the tradeoff between <u>Part A</u> vs <u>Skipping + Part B</u> at runtime

Adaptive Tsallis-INF (AdaTINF) – Unknown σ, α

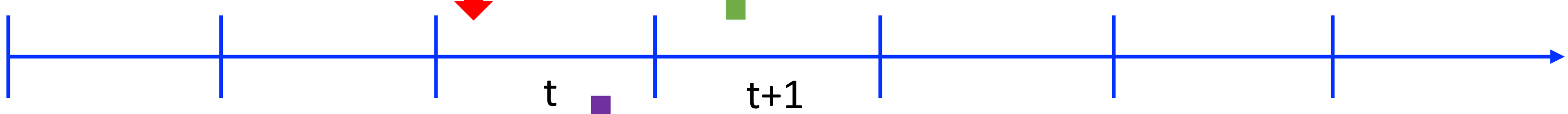
AdaTINF Idea: Using *doubling trick* to tune learning rate and skipping threshold

3. Form x_{t+1} with **FTRL**:

$$x_{t+1} = \arg\min_{x \in \Delta_{[K]}} \left(\eta_{t+1} \left\langle \sum_{s=1}^{t} \hat{l}_s , x \right\rangle + \Psi(x) \right)$$

where $\Psi(x) := -2 \sum_{i=1}^{K} x_i^{1/2}$ (1/2-Tsallis entropy)

1. Pull arm a with prob. $x_{t,a}$ at t



2. Receive l_{t,A_t}, form $\hat{l}_{t,a}$ with **importance sampling** and **skipping**:

$$\hat{l}_{t,a} := \mathbb{I}\left[A_t = a, \left|l_{t,A_t}\right| \le r_t\right] l_{t,A_t} / x_{t,A_t}$$

If total Part A is smaller, double threshold r_t

Part B and Skipping gap
vs
Upper bound of Part A

Adaptive Tsallis-INF (AdaTINF) – Unknown σ, α

Theorem (Informal) AdaTINF achieves the following for the adversarial environment:

$$R_T \leq O\left(\sigma K^{1-\frac{1}{\alpha}} T^{\frac{1}{\alpha}}\right)$$

Remarks

- **Minimax optimal:** Matches the lower bound (Bubeck et al 2013)
- Prior concentration methods heavily rely on knowing α
- Achieving instance-dependent optimality and BoBW are still open

Conclusions

Three novel algorithms for heavy-tailed MAB

- Heavy-Tail Tsallis-INF (HTINF) – known σ, α
  - First to achieve *best-of-both-worlds* for heavy-tailed MAB
  - Applicable to unknown σ, α case (OptTINF): $O(\log T)$ for stochastic and $O(K^{\frac{\alpha-1}{2}} T^{\frac{3-\alpha}{2}})$ for adversarial
- Adaptive Tsallis-INF (AdaTINF) – zero knowledge
  - Optimal $O\left(\sigma K^{1-\frac{1}{\alpha}} T^{\frac{1}{\alpha}}\right)$ regret for adversarial

Reference: J. Huang, Y. Dai, L. Huang, “Adaptive Best-of-Both-Worlds Algorithm for Heavy-Tailed Multi-Armed Bandits,” ICML 2022.

Thank you!

More info: https://people.iiis.tsinghua.edu.cn/~huang