

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

Longbo Huang Tsinghua University

2022 MSRA Data-Driven Optimization Workshop ⁰

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")
- $\{l_{t,a}\}\colon T\times K$ 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")
- $\{l_{t,a}\}\colon T\times K$ loss matrix
- Adversary picks (α, σ) -heavy-tailed distributions $v_{t,1}, ..., v_{t,K}$ 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 $l_{t,A_t} \sim v_{t,A_t}$
- 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")
- $\{l_{t,a}\}\colon T\times K$ loss matrix
- Adversary picks (α, σ) -heavy-tailed distributions $v_{t,1}, ..., v_{t,K}$ with $\mathbb{E}_{X \sim \nu} [|X|^{\alpha}] \leq \sigma^{\alpha}, 1 < \alpha \leq 2$

Natural generalization in both directions

- Stochastic heavy-tailed MAB: $v_{1,a} = v_{2,a} = \cdots v_{T,a}$
- Classical Adversarial MAB: $v_{t,a}$ 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")
- $\{l_{t,a}\}\colon T\times K$ loss matrix
- Adversary picks (α, σ) -heavy-tailed distributions $v_{t,1}, ..., v_{t,K}$ 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

this paper shows that it is indeed to know a ϵ Need to know σ , α before-hand 6

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): $O(logT)$ for stochastic and $O(K)$ α - 1 $\overline{2}$ T $3 - \alpha$ \overline{z}) for adversarial
- Adaptive Tsallis-INF (AdaTINF) zero knowledge
	- Optimal $O\left(\sigma K^{1-\tfrac{1}{\alpha}} T\right)$ $\mathbf 1$ $\overline{\alpha}$) 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

Heavy-Tail Tsallis-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 $i_1, i_2, \dots, i_T \in [K]$

- 1: for $t = 1, 2, \cdots$ do
2: Calculate policy
- 2: Calculate policy with learning rate $\eta_{t}^{-1} = \sigma t^{1/\alpha}$; Pick the regularizer $\Psi(x) = -\alpha \sum_{i=1}^K x_i^{1/\alpha}$ $\longleftarrow 1/\alpha$ -Tsallis entropy $x_t \leftarrow \operatornamewithlimits{argmin}_{x \in \triangle_{[K]}}$ $\sqrt{ }$ $\eta_t \sum$ $t-1$ $s=1$ $\langle \hat{\ell}_s, x \rangle + \Psi(x)$ \$ ropy and an analysis of \overline{C} *2. In stochastically constrained adversarial environments with a unique optimal arm* i Follow-the-Regularized-Leader
- 3: Sample new action $i_t \sim x_t$.
- 4: Calculate the skipping threshold $r_t \leftarrow \Theta_\alpha \eta_t^{-1} x_{t,i_t}^{1/\alpha}$ $\frac{t,i_t}{2}$ overly large los where $\Theta_{\alpha} = \min\{1 - 2^{-\frac{\alpha-1}{2\alpha-1}}, (2-\frac{2}{\alpha})^{\frac{-1}{2-\alpha}}\}.$ ١O ισια (ανόια
s in estimation) Skipping threshold (avoid overly large loss in estimation)
- 5: Play according to i_t and observe loss feedback ℓ_{t,i_t} .
- 6: if $|\ell_{t,i_t}| > r_t$ then
- 7: $\hat{\ell}_t \leftarrow 0.$
8: **else**
- 8: else

9: Construct weighted importance sampling loss estimator $\hat{\ell}_{t,i} \leftarrow \frac{\ell_{t,i}}{x_{t,i}} \mathbb{1}[i = i_t], \forall i \in [K].$ $10:$ end if 11: end for known 1 still guarantees Communications of the 2 state of 2 state o ased due to skipping) and O(log T) instanceregret upper-bound for stochastic instances. Importance Sampling (biased due to skipping)

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

Theorem (Informal) HTINF achieves

• Adversarial environment

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

• Stochastic environment

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

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

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

• Stochastic environment $R_T \leq 0 \vert \sigma$ 2α $\overline{\alpha-1}$ > $a \neq a^*$ Δ _i $\frac{3-\alpha}{\alpha-1}$ $\sqrt{\frac{\alpha-1}{n}} \log T$

Remarks

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

HTINF Regret Analysis $\frac{1}{100}$. For any the set of $\frac{1}{100}$ $\overline{\mathbf{H}}$ $T_{\rm eff}$ can instance with un-NF Regret Analysis **For the** *one-hote* w **and one-hote one-hote** *one-hote* **eivideologie eigenvector and the** *Fore* **eigenvector was expected to any term of the** *separate* **was expected to any term of the** *separate* **was expected to** live Regiel Andiysis **HTINF Regret Analys**

Regret decomposition of HTINF let decomposition of HTINF⊤
tet decomposition of HTINF ret decomposition of filmer "no-regret" performance and O(log T) instance-dependent Regret decomposition of HTINF In this section, we sketch the analysis of Algorithm 1. By

$$
\mathcal{R}_T(y) \triangleq \sum_{t=1}^T \mathbb{E} \left[\langle x_t - y, \mu_t \rangle \right] \quad (y \in \triangle_{[K]})
$$
\n
$$
= \mathbb{E} \left[\sum_{t=1}^T \langle x_t - y, \mu_t - \mu'_t \rangle \right] + \mathbb{E} \left[\sum_{t=1}^T \langle x_t - y, \hat{\ell}_t \rangle \right]
$$
\nSkipping gap

\nFTRL Error

where $\mu'_{t,i} \triangleq \mathbb{E}[\ell_{t,i}1\mathbb{I}[|\ell_{t,i}| \leq r_t] \mid \mathcal{F}_{t-1}, i_t = i]$ is the clipped first part the *skipping gap*, and the second, the *FTRL error*. $f \wedge \pi f$ *e*iⁿ f \wedge f re $\mu'_{t,i} \triangleq \mathbb{E}[\ell_{t,i}1][\ell_{t,i}|\leq r_t] \mid \mathcal{F}_{t-1}, i_t=i]$ is the clipped expectation $\begin{array}{ccccccccccccccccc} \wedge & -\mathbb{E} & \wedge & \wedge & \mathbb{E} & \mathbb{E} & \wedge & \mathbb{E} & \mathbb$ $\mathcal{H}_{t,i} = \mathbb{E}[\mathcal{H}_{t,i} \mathbb{1} | \mathcal{H}_{t,i}] \geq \mathcal{H}_{t} \mid \mathcal{F}_{t-1}, \mathcal{H}_{t} = \mathcal{H} \mid \mathsf{IS} \text{ the G}$

In the following sections, we will show that both $T_{\rm eff}$ wing throchold imposts the regre Skipping threshold impacts the regret # t=1 the rوshold impacts the $\mathbf{P}^{\mathbf{1}}$ $\ddot{}$

- 1.1% HTML (Algorithm 1) with hyper-parameters α larger r_t leads to a smaller skipping gap but a larg when they are known before-hand. When the *distributions'* \mathbf{r} , $\mathop{\mathsf{tipping}}$ gap but a larg • A larger r_t leads to a smaller skipping gap but a larger FTRL error first part the *skipping gap*, and the second, the *FTRL error*. ger r_t leads to a smaller skipping gap but $\overline{}$
- stimes to the deeff-echieved state $O(n^{-1}n^{1/\alpha})$ Julia Gaueon achieved at $r_t = \sigma(\eta_t - \lambda_{t,i_t})$ *heavy-tail parameters* α, σ are both unknown to the agent, t=1 t=1 • Optimal tradeoff achieved at $r_t = \Theta(\eta_t^{-1} x_{t,i_t}^{1/\alpha})$ parts can be controlled and the controlled and the controlled and the controlled \mathcal{L} is the controlled into express- $1/\alpha$

HTINF Regret Analysis $\frac{1}{100}$. For any the set of $\frac{1}{100}$ $\overline{\mathbf{H}}$ $T_{\rm eff}$ can instance with un-NF Regret Analysis **For the** *one-hote* w **and one-hote one-hote** *one-hote* **eivideologie eigenvector and the** *Fore* **eigenvector was expected to any term of the** *separate* **was expected to any term of the** *separate* **was expected to** live Regiel Andiysis **HTINF Regret Analys**

Regret decomposition of HTINF let decomposition of HTINF⊤
tet decomposition of HTINF ret decomposition of filmer "no-regret" performance and O(log T) instance-dependent Regret decomposition of HTINF In this section, we sketch the analysis of Algorithm 1. By

$$
\mathcal{R}_T(y) \triangleq \sum_{t=1}^T \mathbb{E} \left[\langle x_t - y, \mu_t \rangle \right] \quad (y \in \triangle_{[K]})
$$
\n
$$
= \mathbb{E} \left[\sum_{t=1}^T \langle x_t - y, \mu_t - \mu'_t \rangle \right] + \mathbb{E} \left[\sum_{t=1}^T \langle x_t - y, \hat{\ell}_t \rangle \right]
$$
\nSkipping gap

\nFTRL Error

where $\mu'_{t,i} \triangleq \mathbb{E}[\ell_{t,i}1\mathbb{I}[|\ell_{t,i}| \leq r_t] \mid \mathcal{F}_{t-1}, i_t = i]$ is the clipped first part the *skipping gap*, and the second, the *FTRL error*. I , we will show the following sections, we will show that both both both both \mathcal{I} \mathcal{L} $f \wedge \pi f$ *e*iⁿ f \wedge f re $\mu'_{t,i} \triangleq \mathbb{E}[\ell_{t,i}1][\ell_{t,i}|\leq r_t] \mid \mathcal{F}_{t-1}, i_t=i]$ is the clipped expectation we decompose $\frac{1}{2}$ Analysis idea $\begin{array}{ccccccccccccccccc} \wedge & -\mathbb{E} & \wedge & \wedge & \mathbb{E} & \mathbb{E} & \wedge & \mathbb{E} & \mathbb$ $\mathcal{H}_{t,i} = \mathbb{E}[\mathcal{H}_{t,i} \mathbb{1} | \mathcal{H}_{t,i}] \geq \mathcal{H}_{t} \mid \mathcal{F}_{t-1}, \mathcal{H}_{t} = \mathcal{H} \mid \mathsf{IS} \text{ the G}$ # $\frac{1}{2}$ \$x^t − y, µ^t − µ

- 1.1% HTML (Algorithm 1) with hyper-parameters α elf-bounding property" similar to (Zimmert & Sele when they are known before-hand. When the *distributions'* g property" simila dl $o($ $\frac{1}{2}$ nmert & Seld <u>u</u> , • "Self-bounding property" similar to (Zimmert & Seldin, 2019) first part the *skipping gap*, and the second, the *FTRL error*. -bounding property" similar to (Zimmert &
- $\frac{1}{\sqrt{2}}$ result for $\frac{1}{\alpha}$ -Teallis α $<$ $\frac{2}{\alpha}$ regularized FTRL $\frac{1}{2}$ is defined before $\frac{1}{2}$ in the self-of-both $\frac{1}{2}$ is a set-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both-of-both*heavy-tail parameters* α, σ are both unknown to the agent, $1/\alpha$ Teallie $(\alpha \geq 2)$ roquieri \cdots • *First* result for $1/\alpha$ -Tsallis $(\alpha < 2)$ regularized FTRL

HTINF Regret Analysis for the *one-hotel* wild vector α . For any term α for the *one-hot* vector ^y ! ^ei[∗] . For any ^t [∈] [^T], i [∈] [K], **INF Regret Analysis** \cdots \cdots 5.25 To Control the FTRL Errors $\frac{1}{2}$ $\mathbf{F}_{\mathbf{F}}$ the FTRL error part, we follow the regular analysis for $\mathbf{F}_{\mathbf{F}}$ **HILINT NESIEL AND ONLY STATE** $\Delta \Gamma$ Best-of-Algorithms for Γ ≤ 2σ \blacksquare α ret Analy \mathbf{y}

Regret decomposition of HTINF let it is the union of the interest when the set of the set of the interest of the interest of the interest of
Moreover, it is the set of the interest of the \overline{n} pu. $\ddot{}$ & \mathbf{F}_{max} algorithms. Note that our skipping mechanism is skipping mechanism is skipping mechanism is skipping mechanism is skipping mechanism in the skipping mechanism is seen in the skipping mechanism is seen in the Regret decomposition of HTINF The skipping gap participant in the skipping gap Γ

HTINF Regret Analysis for the *one-hotel* wild vector α . For any term α for the *one-hot* vector ^y ! ^ei[∗] . For any ^t [∈] [^T], i [∈] [K], **INF Regret Analysis** \cdots \cdots 5.25 To Control the FTRL Errors $\frac{1}{2}$ $\mathbf{F}_{\mathbf{F}}$ the FTRL error part, we follow the regular analysis for $\mathbf{F}_{\mathbf{F}}$ \ddotsc $t_{\rm max} = E \left(\frac{1}{1 + \frac{1}{1$ JF Regret Analys t**Γ** Decise Les **all**

Regret decomposition of HTINF let it is the union of the interest when the set of the set of the interest of the interest of the interest of
Moreover, it is the set of the interest of the $\boldsymbol{\beta}$ ct accorded by $\boldsymbol{\beta}$ \overline{n} pu. $\ddot{}$ \$x^t − y, µ^t − µ & \mathbf{F}_{max} algorithms. Note that our skipping mechanism is skipping mechanism is skipping mechanism is skipping mechanism is skipping mechanism in the skipping mechanism is seen in the skipping mechanism is seen in the Regret decomposition of HTINF [≤] ^σ^αr¹−^α ≤ 2σ i mposition 1/α
1/α−1 · α−1

$\mathsf{u}\mathsf{a}\mathsf{p}\mathsf{u}\mathsf{v}\mathsf{e}$ is anis-hvr (Add Hivr) = Unknown for the *one-hot* vector ^y ! ^ei[∗] . For any ^t [∈] [^T], i [∈] [K], aptive Tsallis-INF (AdaTINF) – Un we decree to different two parts Adaptive Tsallis-INF (AdaTINF) – Unknown σ , α \blacksquare . ₁ $\ddot{ }$ σ To To The FTRL THE TERRITAL TEACHER TEAL TEACHER FTRL error part, we follow the FTRL error part, we follow the regular analysis for \mathcal{F} We can bound the expectation of each summand in part (B) n known o , α

daTINF Idea: adjust the tradeoff between Part A vs Skipping + Part B I funtime sections, we will show that both I $\frac{1}{\sqrt{1-\frac{1$ ρ (Simmert α) from 1/2-Tsallis en-Seldin, 2019) from 1/2-Tsallis en-Seldin, 2019) from 1/2-Tsallis en-Seldin, 2019 16 t DΨ
CW DW DW DW DW $\overline{}$, $\overline{}$ at runtime $\frac{10}{\pi}$ AdaTINF Idea: adjust the tradeoff between Part A vs Skipping + Part B

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

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

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

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

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

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): $O(logT)$ for stochastic and $O(K^{\overline{-2}}T^{\overline{-2}})$ for adversarial α - 1 $3 - \alpha$
- Adaptive Tsallis-INF (AdaTINF) zero knowledge

\n- Optimal
$$
O\left(\sigma K^{1-\frac{1}{\alpha}}T^{\frac{1}{\alpha}}\right)
$$
 regret for adversarial
\n

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

Institute for Interdisciplinary Information Sciences, Tsinghua University

Thank you!

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