Towards Fully Adaptive Regret Minimization in Heavy-Tailed Bandits

Gianmarco Genalti, Lupo Marsigli, Nicola Gatti and Alberto Maria Metelli
{gianmarco.genalti, nicola.gatti, albertomaria.metelli}@polimi.it, lupo.marsigli@mail.polimi.it

- In stochastic multi-armed bandits, at each round $t \in \{1, \dots, T\}$ an agent choose among K (unknowdistributions $\{\nu_i\}_{i \in \{1,\dots,K\}}$ and observes a sample reward X_t. The goal is to **minimize the expected cumulative regret** w.r.t. the best action.

- **Stochastic heavy-tailed bandits** gained popularity over the last years, extending the framework from sub-gaussian distributions to scenarios with (possibly) infinite variance, i.e.

$$\mathbb{E}_{\nu_i}[|X|^{1+\epsilon}] \leq u, \quad \epsilon \in (0,1], \quad \forall i \in \{1, \dots, K\},$$

all moments of order $> 1 + \epsilon$ are non-finite.

- **Most of the literature assumes both ϵ and u to be known** to the agent, but in practice this is usually a hard requirement to satisfy. We study the *adaptive heavy-tailed bandit problem*, in which the learner has no knowledge on these quantities.

- We show that, in general, **attaining optimal performance while being adaptive w.r.t. to either ϵ or u is impossible**. However, under a specific assumption, **our algorithm is capable of matching the best possible performance of the setting**.