A Regret Lower Bound for u-Adaptive Heavy-Tailed Bandits

Gianmarco Genalti
Alberto Maria Metelli

Politecnico di Milano

Abstract

The heavy-tailed bandit problem, introduced by Bubeck et al. [2013], is a variant of the stochastic multi-armed bandit problem where the reward distributions have finite absolute raw moments of maximum order $1+\epsilon$, uniformly bounded by a constant $u < +\infty$, for some $\epsilon \in (0, 1]$. In this technical note, we provide a lower bound for the regret of every algorithm that *adapts* to u, *i.e.*, is unaware of the value of u or of any upper bound of it. Our bound closely follows the style of the one proposed in Hadiji and Stoltz [2020], and exposes a trade-off between the instance-dependent and the worst-case rates.

1 Preliminaries

We recall some fundamental notions on heavy-tailed bandits and the required notations for regret rates defined in Hadiji and Stoltz [2020].

In the stochastic multi-armed bandit problem (MAB), a learner is faced with $K \in \mathbb{N}$ arms repeatedly for $T \in \mathbb{N}$ rounds. Every time an action $i \in [K]$ is selected, a reward X is sampled from the distribution ν_i. We call $\underline{\nu} \coloneqq \{\nu_i\}_{i\in[K]}$ an *instance*. Let $\mathcal{H}_{\epsilon,u}$ be the set of instances such that

$$\mathcal{H}_{\epsilon,u} \coloneqq \{\underline{\nu} : \mathbb{E}_{\nu_i}[|X|^{1+\epsilon}] \leq u \quad \forall \nu_i \in \underline{\nu}\},$$

where $\epsilon \in (0, 1]$ and $u \in \mathbb{R}^+$. We call the bandit problem over the instances defined in this way a heavy-tailed bandit problem. Let $\mu_i \coloneqq \mathbb{E}_{X\sim\nu_i}[X]$ and I_t be the action selected at round $t \in [T]$. Then, the learner's goal is to minimize the expected cumulative regret, defined as:

$$\mathbb{E}[R_T(\underline{\nu})] \coloneqq \mathbb{E}\left[\sum_{t\in[T]} (\mu^* - \mu_{I_t})\right], \qquad \text{where} \qquad \mu^* \coloneqq \max_{i\in[K]} \mu_i.$$

It is customary to provide theoretical guarantees for an algorithm by upper-bounding its expected cumulative regret. Here, we are interested in algorithms that are unaware of u or any other possible information about it, such as an upper bound. This setting is called *adaptive* HTMAB, and in particular u-Adaptive since the adaptation only concerns u. The problem of adaptation in HTMAB recently gained popularity, and we refer to Genalti et al. [2024] for a comprehensive literature review on the problem. There are two main ways to express bounds over the expected cumulative regret.

Definition 1.1 (Moment-free distribution-free regret bounds)**.** A strategy for stochastic, heavy-tailed bandits is adaptive to the unknown moment u of order $1+\epsilon$ with a moment-free distribution-free regret bound $\Phi_{free} : \mathbb{N} \to [0, +\infty)$ if for all real numbers u, the strategy ensures, without the knowledge of u:

$$\forall \underline{\nu} \in \mathcal{H}_{\epsilon,u}, \qquad \forall T \geq 1, \qquad R_T(\underline{\nu}) \leq u^{\frac{1}{1+\epsilon}} \Phi_{free}(T).$$

Definition 1.2 (Distribution-dependent rates for adaptation)**.** A strategy for stochastic, heavy-tailed bandits is adaptive to the unknown centered moment u of order $1+\epsilon$ with a distribution-dependent rate $\Phi_{dep} : \mathbb{N} \to [0, +\infty)$ if for all real numbers u, the strategy ensures, without the knowledge of u:

$$\forall \underline{\nu} \in \mathcal{H}_{\epsilon,u}, \qquad \limsup_{T \to +\infty} \frac{R_T(\underline{\nu})}{\Phi_{dep}(T)} < +\infty.$$

2 Lower Bound

We closely follow the procedure developed in Hadiji and Stoltz [2020], together with the instance construction of Bubeck et al. [2013]. We show that there exists a trade-off between the distribution-free and the distribution-dependent regret bounds, for any algorithm that is adaptive to u.

Theorem 2.1 (Existence of a trade-off)**.** A strategy with scale-free distribution-free rate of $\Phi_{free}(T) = o(T)$ may only achieve distribution-dependent rates $\Phi_{dep}(T)$ for adaptation satisfying:

$$\Phi_{dep}(T) \Phi_{free}(T)^{\frac{1+\epsilon}{\epsilon}} \geq T^{\frac{1+\epsilon}{\epsilon}},$$

more precisely, the regret of such a strategy is lower bounded as follows, $\forall \underline{\nu} \in \mathcal{H}_{\epsilon,u}$:

$$\liminf_{T \to +\infty} \frac{R_T(\underline{\nu})}{(T/\Phi_{free}(T))^{\frac{1+\epsilon}{\epsilon}}} \geq \frac{1}{16} \sum_{i:\Delta_i > 0} \Delta_i.$$

Proof. Consider an instance $\underline{\nu} \in \mathcal{H}_{\epsilon,u}$ s.t. there's at least a suboptimal arm a. For this arm, assume $\mu_a := \mathbb{E}_{\nu_a}[X]$ and $u := \mathbb{E}_{\nu_a}[|X|^{1+\epsilon}]$. Let the suboptimality gap for arm a be Δ_a. For some $\beta \in [0, 1]$, we also consider the alternative instance $\underline{\nu}'$, where all distributions are the same as $\underline{\nu}$ except for $\nu'_a = (1-\beta)\nu_a + \beta \delta_{\mu_a + 2\Delta_a \beta^{-1}}$. We have that $\mathbb{E}_{\underline{\nu}'}[X] = \mu_a + 2\Delta_a$ and $u' := \mathbb{E}_{\underline{\nu}'}[|X|^{1+\epsilon}] = (1-\beta)u + \beta \left(\mu_a + \frac{2\Delta_a}{\beta}\right)^{1+\epsilon}$. Note that $\Delta'_a = \Delta_a$. We also compute $KL(\nu_a || \nu'_a) = \ln\left(\frac{1}{1-\beta}\right)$.

For $\beta < \frac{1}{2}$, we have that $\ln\left(\frac{1}{1-\beta}\right) < 2\beta \ln 2$.

Moreover

$$\begin{aligned} KL(p,q) &\geq \underbrace{p \ln p + (1-p)\ln(1-p)}_{\geq -\ln 2} + \underbrace{p \ln \frac{1}{q}}_{\geq 0} + (1-p)\ln\frac{1}{1-q} \\ &\geq (1-p)\ln\left(\frac{1}{1-q}\right) - \ln 2 \quad \forall p, q \in [0,1]. \end{aligned}$$

We use these inequalities together with the fundamental fact that

$$KL\left(\frac{\mathbb{E}_{\underline{\nu}}[N_a(T)]}{T}, \frac{\mathbb{E}_{\underline{\nu}'}[N_a(T)]}{T}\right) \leq \mathbb{E}_{\underline{\nu}}[N_a(T)] \ln\left(\frac{1}{1-\beta}\right)$$

to obtain the following:

$$\left(1 - \frac{\mathbb{E}_{\underline{\nu}}[N_a(T)]}{T}\right) \ln\left(\frac{1}{1 - \mathbb{E}_{\underline{\nu}'}[N_a(T)]/T}\right) - \ln 2 \leq (2\beta \ln 2)\mathbb{E}_{\underline{\nu}}[N_a(T)]. \tag{1}$$

Now it's time to use Definition 1.1:

$$\Delta_a \mathbb{E}_{\underline{\nu}}[N_a(T)] \leq R_T(\underline{\nu}) \leq u^{\frac{1}{1+\epsilon}} \Phi_{free}(T), \tag{2}$$

similarly, it holds

$$\Delta_a(T - \mathbb{E}_{\underline{\nu}'}[N_a(T)]) = \Delta_a'(T - \mathbb{E}_{\underline{\nu}'}[N_a(T)]) \leq R_T(\underline{\nu}') \leq (u')^{\frac{1}{1+\epsilon}} \Phi_{free}(T). \tag{3}$$

We now plug Eq.(2)-(3) in Eq.(1):

$$\left(1 - \frac{u^{\frac{1}{1+\epsilon}} \Phi_{free}(T)}{T\Delta_a}\right) \ln\left(\frac{T\Delta_a}{(u')^{\frac{1}{1+\epsilon}} \Phi_{free}(T)}\right) - \ln 2 \leq (2\beta \ln 2)\mathbb{E}_{\underline{\nu}}[N_a(T)]. \tag{4}$$

We now take $\beta = \beta_T = \alpha^{-1} \left(\frac{\Phi_{free}(T)}{T}\right)^{\frac{1+\epsilon}{\epsilon}}$ for some constant $\alpha > 0$. By assumption, $\Phi_{free}(T) = o(T)$, so $\beta_T \to 0$ if $T \to +\infty$.

Substituting in the definition of u', and taking the limit, yields:

$$\begin{aligned}
\liminf_{T\to\infty} u_T' &= \liminf_{T\to\infty} \left((1-\beta_T)u + \beta_T(\mu_a + 2\Delta_a \beta_T^{-1})^{1+\epsilon}\right) \\
&= \liminf_{T\to\infty} \left(u + \beta_T \left(\mu_a^{1+\epsilon} + (2\Delta_a \beta_T^{-1})^{1+\epsilon} - u\right)\right) \\
&= \liminf_{T\to\infty} \left(u + (2\Delta_a)^{1+\epsilon} \beta_T^{-\epsilon}\right) \\
&= \liminf_{T\to\infty} \left(u + \alpha^\epsilon (2\Delta_a)^{1+\epsilon} \left(\frac{T}{\Phi_{free}(T)}\right)^{1+\epsilon}\right) \\
&= \liminf_{T\to\infty} \alpha^\epsilon (2\Delta_a)^{1+\epsilon} \left(\frac{T}{\Phi_{free}(T)}\right)^{1+\epsilon},
\end{aligned}$$

which implies

$$\liminf_{T\to\infty} (u_T')^{\frac{1}{1+\epsilon}} \Phi_{free}(T) = \alpha^{\frac{\epsilon}{1+\epsilon}} 2\Delta_a T.$$

Substituting in the LHS of Eq.(4), we get:

$$\begin{aligned}
&\liminf_{T\to\infty} \left(1 - \frac{u^{\frac{1}{1+\epsilon}} \Phi_{free}(T)}{T\Delta_a}\right) \ln\left(\frac{T\Delta_a}{(u_T')^{\frac{1}{1+\epsilon}} \Phi_{free}(T)}\right) - \ln 2 = \\
&= \liminf_{T\to\infty} \ln\left(\frac{T\Delta_a}{\alpha^{\frac{\epsilon}{1+\epsilon}} 2\Delta_a T}\right) - \ln 2 \\
&= \ln\left(\frac{1}{4\alpha^{\frac{\epsilon}{1+\epsilon}}}\right).
\end{aligned}$$

We now choose $\alpha = 8^{-\frac{1+\epsilon}{\epsilon}}$, and by Equation (4) we get:

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}[N_a(T)]}}{(T/\Phi_{free}(T))^{\frac{1+\epsilon}{\epsilon}}} \geq \frac{\alpha}{2\ln 2} \ln\left(\frac{1}{4\alpha^{\frac{\epsilon}{1+\epsilon}}}\right) \geq \frac{1}{16}. \tag{5}$$

Using regret decomposition, Equation (5) yields a relationship involving the regret in instance $\underline{\nu}$:

$$\liminf_{T\to+\infty} \frac{R_T(\underline{\nu})}{(T/\Phi_{free}(T))^{\frac{1+\epsilon}{\epsilon}}} \geq \frac{1}{16} \sum_{i:\Delta_i>0} \Delta_i. \tag{6}$$

Using Definition 1.2, we also get

$$\Phi_{dep}(T)\Phi_{free}(T)^{\frac{1+\epsilon}{\epsilon}} \geq T^{\frac{1+\epsilon}{\epsilon}},$$

which concludes the proof. ∎

If we impose the two rates to be equal, *i.e.* $\Phi_{dep} = \Phi_{free}$, we get that $\Phi_{free}(T) \geq \Omega\left(T^{\frac{1+\epsilon}{1+2\epsilon}}\right)$. When $\epsilon = 1$, we have $\Omega\left(T^{\frac{2}{3}}\right)$, which is higher than the $\Omega\left(\sqrt{T}\right)$ lower bound obtained in bounded range bandits.

References

S. Bubeck, N. Cesa-Bianchi, and G. Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013.

G. Genalti, L. Marsigli, N. Gatti, and A. M. Metelli. (ε, u)-adaptive regret minimization in heavy-tailed bandits. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 1882–1915. PMLR, 2024.

H. Hadiji and G. Stoltz. Adaptation to the range in k-armed bandits. *arXiv preprint arXiv:2006.03378*, 2020.