Algorithms for Infinitely Many-Armed Bandit (Supplementary file)

Yizao Wang
INRIA Futurs Lille, SequeL project,
40 avenue Halley, 59650 Villeneuve d'Ascq, France
yizwang@umich.edu

Jean-Yves Audibert
Université Paris Est, Ecole des Ponts, ParisTech, Certis
& Willow - ENS / INRIA, Paris, France
audibert@certis.enpc.fr

Rémi Munos
INRIA Futurs Lille, SequeL project,
40 avenue Halley, 59650 Villeneuve d'Ascq, France
remi.munos@inria.fr

Theorem 3 Any algorithm suffers a regret larger than $cn^{\frac{\beta}{1+\beta}}$ for some small enough constant c depending on c_2 and β.

Proof of Theorem 3. An elementary event of the probability space is characterized by the infinite sequence $I_1, I_2, \dots$ of arms and by the infinite sequences of rewards corresponding to each of the arm: $X_{I_1,1}, X_{I_1,2}, \dots, X_{I_2,1}, X_{I_2,1}, \dots$, and so on. Arm I_1 is the first arm drawn, $I_2 \neq I_1$ is the second one, and so on. Let $0 < \delta < \delta' < \mu^*$. Let K^* denote the smallest ℓ such that $\mu_{I_\ell} > \mu^* - \delta$. Let $\bar{K}$ be the number of arms in $\{I_1, \dots, I_{K^*-1}\}$ with expected reward smaller than or equal to $\mu^* - \delta'$. An algorithm will request a number of arms K, which is a random variable (possibly depending on the obtained rewards). Let $\hat{\mu}$ be the expected reward of the best arm in $\{I_1, \dots, I_K\}$. Let $\kappa > 0$ a parameter to be chosen. We have

$$\begin{aligned} R_n &= R_n \mathbf{1}_{\hat{\mu} \le \mu^* - \delta} + R_n \mathbf{1}_{\hat{\mu} > \mu^* - \delta} \\ &\ge n\delta \mathbf{1}_{\hat{\mu} \le \mu^* - \delta} + \bar{K}\delta' \mathbf{1}_{\hat{\mu} > \mu^* - \delta} \\ &\ge n\delta \mathbf{1}_{\hat{\mu} \le \mu^* - \delta} + \kappa\delta' \mathbf{1}_{\hat{\mu} > \mu^* - \delta; \bar{K} \ge \kappa}, \end{aligned}$$

where the first inequality uses that $\hat{\mu} > \mu^* - \delta$ implies that the arms $I_1, \dots, I_{K^*}$ have been at least tried once. By taking expectations on both sides and taking $\kappa = n\delta/\delta'$, we get

$$\mathbb{E}R_n \ge n\delta\mathbb{P}(\hat{\mu} \le \mu^* - \delta) + \kappa\delta'\big(\mathbb{P}(\hat{\mu} > \mu^* - \delta) - \mathbb{P}(\bar{K} < \kappa)\big) = \delta'\kappa\mathbb{P}(\bar{K} \ge \kappa).$$

Now the random variable $\bar{K}$ follows a geometric distribution with parameter $p = \frac{\mathbb{P}(\mu > \mu^* - \delta)}{\mathbb{P}(\mu \notin (\mu^* - \delta', \mu^* - \delta])}$. So we have $\mathbb{E}R_n \ge \delta'\kappa(1-p)^\kappa$. Taking $\delta = \delta' n^{-1/(\beta+1)}$ and δ' a constant value in $(0, \mu^*)$ (for instance $(2c_2)^{-1/\beta}$ to ensure $p \le 2c_2\delta^\beta$), we have $\kappa = n^{\frac{\beta}{1+\beta}}$ and p is of order $1/\kappa$ and obtain the desired result.

Theorem 4 For any horizon time $n \ge 2$, the expected regret of the UCB-AIR algorithm satisfies

$$\mathbb{E}R_n \le \begin{cases} C(\log n)^2\sqrt{n} & \text{if } \beta < 1 \text{ and } \mu^* < 1 \\ C(\log n)^2 n^{\frac{\beta}{1+\beta}} & \text{otherwise, i.e. if } \mu^* = 1 \text{ or } \beta \ge 1 \end{cases} \tag{1}$$

with C a constant depending only on c_1, c_2 and β.

Proof of Theorem 4. We essentially need to adapt the proof of Theorem 1. We recall that K_n denote the number of arms played up to time n. Let $I_1, \ldots, I_{K_n}$ denote the selected arms: I_1 is the first arm drawn, I_2 the second, and so on. Let S_k denote the time arm k being played for the first time. $1 = S_{I_1} < S_{I_2} < \cdots < S_{I_{K_n}}$. Since arms $I_1, \ldots, I_{K_n}$ progressively enter in competition, Lemma 1 no longer holds but an easy adaptation of its proof shows that for $k \in \{I_1, \ldots, I_{K_n}\}$,

$$\mathbb{E}(T_k(n)|I_1, \ldots, I_{K_n}) \leq u + \sum_{t=u+1}^{n} \sum_{s=u}^{t} \mathbb{P}\big(B_{k,s,t} > \tau\big) + \Omega_k \tag{2}$$

with

$$\Omega_k = \sum_{t=u+1}^{n} \prod_{k' \neq k, S_{k'} \leq t} \mathbb{P}\big(\exists s' \in [0,t],\ B_{k',s',t} \leq \tau\big)\,.$$

As in the proof of Theorem 1, since the exploration sequence satisfies $\mathcal{E}_t \geq 2\log(10\log t)$, we have $\mathbb{P}\big(\exists s' \in [0,t],\ B_{k',s',t} \leq \tau\big) \leq 1/2$ for arms k' such that $\mu_{k'} \geq \tau$. Consequently, letting $N_{\tau,k,t}$ denote the cardinal of the set $\{k' : k' \neq k, \mu_{k'} \geq \tau, S_{k'} \leq t\}$, we have

$$\Omega_k \leq \sum_{t=1}^{n} 2^{-N_{\tau,k,t}}\,.$$

Let us first consider the case $\mu^* = 1$ or $\beta \geq 1$. In the case of UCB-AIR, S_{I_j} is the smallest integer strictly larger than $(j-1)^{(\beta+1)/\beta}$. To shorten notation, let us write S_j for S_{I_j}. According to the arm-increasing rule (try a new arm if $K_{t-1} < t^{\beta/(\beta+1)}$), $[S_j, S_{j+1})$ is the time interval in which the competing arms are $I_1, I_2, \ldots, I_j$.

As in the proof of Theorem 1, we consider $\tau = \mu^* - \Delta_k/2$. We have

$$\begin{aligned}
\mathbb{E}(\Omega_{I_\ell}|I_\ell = k) &\leq \sum_{j=1}^{K_n} \sum_{t=S_j}^{S_{j+1}-1} \mathbb{E}\Big(2^{-N_{\tau,k,S_j}}|I_\ell = k\Big) \\
&= \sum_{j=1}^{K_n} (S_{j+1} - S_j)\mathbb{E}\Big(2^{-N_{\tau,k,S_j}}|I_\ell = k\Big) \\
&\leq \sum_{j=1}^{K_n} (S_{j+1} - S_j)\mathbb{E}\Big(2^{-N_{\tau,\infty,S_{j-1}}}\Big)\,.
\end{aligned} \tag{3}$$

Since $N_{\tau,\infty,S_{j-1}}$ follows a binomial distribution with parameter $j-1$ and $\mathbb{P}(\mu \geq \tau)$, we have

$$\mathbb{E}\Big(2^{-N_{\tau,\infty,S_{j-1}}}\Big) = (1 - \mathbb{P}(\mu \geq \tau)/2)^{j-1},$$

and

$$\begin{aligned}
\sum_{j=1}^{K_n} (S_{j+1} - S_j)\mathbb{E}\Big(2^{-N_{\tau,\infty,S_{j-1}}}\Big) &= \sum_{j=1}^{K_n} \big(S_{j+1} - S_j\big)(1 - \mathbb{P}(\mu \geq \tau)/2)^{j-1} \\
&\leq \sum_{j=1}^{K_n} (1 + \tfrac{\beta+1}{\beta} j^{\frac{1}{\beta}})(1 - \tilde{c}[2(\mu^* - \tau)]^{\beta})^{j-1},
\end{aligned} \tag{4}$$

where $\tilde{c} = c_1 2^{-1-\beta}$. Plugging (4) into (3), we obtain

$$\mathbb{E}(\Delta_{I_\ell}\Omega_{I_\ell}) \leq \tfrac{2\beta+1}{\beta} \sum_{j=1}^{K_n} j^{\frac{1}{\beta}} \mathbb{E}\big(\Delta_{I_\ell}\big[1 - \tilde{c}\Delta_{I_\ell}^{\beta}\big]^{j-1}\big).$$

Now this last expectation can be bounded by the same computations as for $\mathbb{E}\chi(\Delta_1)$ in the proof of Theorem 1. We have, for appropriate positive constants C_1 and C_2 depending on c_1 and β,

$$\mathbb{E}(\Delta_{I_\ell}\Omega_{I_\ell}) \leq C_1 \sum_{j=1}^{K_n} j^{\frac{1}{\beta}} j^{-\frac{1}{\beta}} \tfrac{\log j}{j} \leq C_2(\log K_n)^2\,. \tag{5}$$

Using (2) and $\mathbb{E}R_n = \sum_{\ell=1}^{K_n} \mathbb{E}(\Delta_{I_\ell}\Omega_{I_\ell})$, we obtain

$$\mathbb{E}R_n \leq K_n \mathbb{E}\bigg\{\Big[50\Big(\tfrac{V(\Delta_1)}{\Delta_1} + 1\Big)\log n\Big] \wedge (n\Delta_1) + C_2(\log K_n)^2\bigg\}, \tag{6}$$

from which Theorem 4 follows for the case $\mu^* = 1$ or $\beta \geq 1$. For the case $\beta < 1$ and $\mu^* < 1$, replacing $\frac{\beta}{\beta+1}$ by $\frac{\beta}{2}$ leads to a similar version of (5) as

$$\mathbb{E}(\Delta_{I_\ell}\Omega_{I_\ell}) \leq C_1 \sum_{j=1}^{K_n} j^{\frac{2}{\beta}-1} j^{-\frac{1}{\beta}} \tfrac{\log j}{j} \leq C_2(\log K_n) K_n^{\frac{1-\beta}{\beta}}\,,$$

which gives the desired convergence rate since K_n is of order $n^{\beta/2}$.