# Noise Free Multi-armed Bandit Game

Atsuyoshi Nakamura[1(✉)], David P. Helmbold[2], and Manfred K. Warmuth[2]

[1] Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo 060-0814, Japan
atsu@main.ist.hokudai.ac.jp
[2] University of California at Santa Cruz, Santa Cruz, USA
{dph,manfred}@cse.ucsc.edu

**Abstract.** We study the loss version of adversarial multi-armed bandit problems with one lossless arm. We show an adversary's strategy that forces any player to suffer $K - 1 - O(1/T)$ loss where $K$ is the number of arms and $T$ is the number of rounds.

**Keywords:** Algorithmic learning · Online learning · Bandit problem

## 1 Introduction

In this paper, we study a kind of the loss version of adversarial multi-armed bandit problems with one lossless arm, that is, the problem in the noise-free case. The gain (reward) version of an adversarial multi-armed bandit problem was studied first by Auer et al. [1] and its loss version has been also studied in some papers such as [2,3]. To the best of our knowledge, however, the problem in the noise-free setting has not been studied yet. This could be because the problem is too trivial to study. In fact, it is easy in the *full-information* case; with $\{0, 1\}$ losses and $K$ arms, loss $\sum_{i=2}^{K}(1/i) = \Theta(\ln K)$ is achieved by the minmax strategy: the adversary's maximization strategy sets the loss of an arm with the highest probability of being chosen to 1 in addition to the past lossy arms, and the player's minimization strategy always chooses one of the arms with no loss so far randomly with equal probability. In the bandit case, however, it does not seem so trivial because the arms with no loss so far in player's observation might already have suffered 1-loss. Thus, the adversary may have to stick to its loss assignments, waiting for the player to choose one of the lossy arms.

In this paper, we focus on an adversary's strategy. The adversary's strategy studied in [1] selects a best arm randomly and sets losses of the non-best arms to 1 with probability 1/2 and sets the loss of the best arm to 1 with probability $1/2 - \epsilon$ at each time for some small $\epsilon$. Their strategy can be modified for the noise free case by changing the probabilities of 1-loss to $\epsilon$ and 0. This adversary's strategy is very weak against the player's strategy that sticks to the same random arm until he/she suffers 1-loss, forcing only $(K - 1)/2$ loss in the $K$-arm case. We show an adversary's strategy that forces any player to suffer $K - 1 - O(1/T)$ loss for $K$-arm and $T$-round case.


A.-H. Dediu et al. (Eds.): LATA 2016, LNCS 9618, pp. 412–423, 2016.
DOI: 10.1007/978-3-319-30000-9_32

## 2 Problem Setting

For any natural numbers $i, j$ with $i \leq j$, $[i..j]$ denotes the set $\{i, \dots, j\}$ and we let $[j]$ denote $[1..j]$. For any sequence $x_1, \dots, x_n$, we let $\mathbf{x}[b..e]$ denote its contiguous subsequence $x_b, \dots, x_e$.

The *noise-free multi-armed bandit problem* we consider here is the loss version of an adversarial multi-armed bandit problem with one lossless arm. It is a $T$-round game between a player and an adversary. There are $K$ arms (of slot machines): arm $1, \dots,$arm $K$. At each time $t = 1, \dots, T$, the adversary picks a loss $\ell_{t,i} \in [0, 1]$ for each arm $i \in [K]$. Let $\boldsymbol{\ell}_t \in [0, 1]^K$ denote a $K$-dimensional vector $(\ell_{t,1}, \dots, \ell_{t,K})$. The player, who does not know $\boldsymbol{\ell}_t$, chooses arm $I_t$ and suffers loss $\ell_{t,I_t}$. The player's and the adversary's objectives are minimization and maximization, respectively, of player's (expected) cumulative loss $\sum_{t=1}^{T} \ell_{t,I_t}$.

The most popular measure for evaluating player's strategies is *regret*, which is difference between player's and the best arm's cumulative losses. Throughout this paper, we assume that there is an arm whose cumulative loss is zero. In this case, regret coincides with cumulative loss. Note that this assumption constrains the adversary's choices.

We allow the player to use a randomized strategy, so at each time $t$ the player's choice $I_t$ is a random variable. Let $i_t$ denote a realization of random variable $I_t$. We call $(i_t, \ell_{t,i_t})$ a player's *observation* at time $t$ and denote it by $o_t$. Each player's choice $I_t$ can depend only on his/her past observations $\mathbf{o}[1..t-1]$. The adversary that we consider here is assumed to have infinite computation power; no limitation is set on adversary's computational time and space. The Adversary is also allowed to behave adaptively: the adversary's decision $\boldsymbol{\ell}_t$ can depend on both the player's past choices $\mathbf{i}[1..t-1]$ and the adversary's past decisions $\boldsymbol{\ell}[1..t-1]$.

*Example 1.* Let $K = 2$. Consider a *randomized consistent conservative player* who chooses each arm $i$ with equal probability at $t = 1$ and continues to choose the same arm $i$ until suffering a non-zero loss, and after that chooses the other arm, which must be a lossless arm by the assumption of one lossless arm. For this player, $\mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] = 1/2$ if $\boldsymbol{\ell}_1 = \cdots = \boldsymbol{\ell}_T = (1, 0)$ or $(0, 1)$. The adversary, however, can achieve $\mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] = 1$ using $\boldsymbol{\ell}_1 = (0, 0)$ because the adversary can know $I_2$ from $I_1$. For a mere *randomized consistent player* who chooses each arm of no loss so far with equal probability, using $\boldsymbol{\ell}_1 = (0, 0)$ only reduces $\mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right]$ and the best strategy of the adversary is to use $\boldsymbol{\ell}_1 = \cdots = \boldsymbol{\ell}_T = (1, 0)$ or $(0, 1)$ which achieves $\mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] = 1 - (1/2)^T$. Note that, for this loss sequence, $\mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] = 1/2$ holds in the full-information setting because the lossless arm can be identified at $t = 1$ regardless of the loss suffered by the player.

**Algorithm 1.** RepeatW&S$[K, T]$

**parameter**: $K$: number of arms, $T$: number of trials
**initialize** : $t_0, \ldots, t_m, k_1, \ldots, k_m \leftarrow$ the solution of Problem 2, $d \leftarrow 1$

```
for time t = 1, ..., T do
    if t ≥ t_{d−1} then
        b ← t_{d−1}, e ← t_d − 1
        c ← BestSwitchingTime(b, e, o[1..b − 1], k_d)
        d ← d + 1
    end
    ℓ_t ← Wait&Sticking[b, c, e, o[1..b − 1], k_d](t, o[b, ..t − 1])
    Observe the player's choice i_t
end
```

**Algorithm 2.** BestSwitchingTime$(b, e, \mathbf{o}[1..b-1], k)$

**input** : $b, e \in \mathbb{N}$: beginning and ending times with $1 \le b \le e$
$\mathbf{o}[1..b-1]$: players observations from time 1 to time $b-1$
$k$: number of no-loss arms to switch to 1-loss
**output**: $c^*$: best time to switch from waiting to sticking

$S \leftarrow$ the set of arms with no loss by time $b-1$
$p_{\max} = -1$
**for** $c = b, \ldots, e$ **do**

$$p_c \leftarrow E_{I[b..e-1]}\left[\max_{s \subseteq S, |s|=k} P\left\{\sum_{t=c}^{e} \ell_{t,I_t} \ge 1 \;\middle|\; \begin{array}{l} \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], I[b..c-1], \\ \boldsymbol{\ell}_b = \cdots = \boldsymbol{\ell}_{c-1} = \mathbb{1}_{[K]\setminus S}, \\ \boldsymbol{\ell}_c = \cdots = \boldsymbol{\ell}_e = \mathbb{1}_{([K]\setminus S)\cup s} \end{array}\right\}\right]$$

  **if** $p_c > p_{max}$ **then**
    $c^* \leftarrow c$
    $p_{\max} \leftarrow p_c$
  **end**
**end**
**return** $c^*$

## 3 Adversary's Strategy

For any set $S \subseteq [K]$, define $\mathbb{1}_S$ to be the $K$-dimensional $\{0,1\}$-vector whose $i$th component is 1 if and only if $i \in S$. At any time $t \in [T]$, the decision $\boldsymbol{\ell}_t$ made by our adversary algorithm RepeatW&S$[K, T]$ (Algorithm 1) is $\mathbb{1}_{S_t}$ for some set $S_t \subseteq [K]$ and will satisfy $\boldsymbol{\ell}_1 \le \cdots \le \boldsymbol{\ell}_T$, that is, once the $i$th component becomes 1, it never becomes 0. Based on the loss analysis in Sect. 4, for some natural number $m \in [K-1]$, the adversary divides $[T]$ into $m$ parts $[t_0..t_1-1], \cdots, [t_{m-1}, t_m-1]$, and also divides $K-1$ into $m$ non-negative integers $k_1, \ldots, k_m$ with $\sum_{i=1}^{m} k_i = K-1$. We explain how to divide $[T]$ and $K-1$ later in this section.

During the $i$th period $[t_{i-1}..t_i - 1]$, that is, for times $t \in [t_{i-1}, t_i - 1]$, the adversary switches the loss vector $\boldsymbol{\ell}_t$ at most once: beginning with $\mathbb{1}_{[K]\setminus S}$, where $S$ is the set of lossless arms so far, it calculates the best time $c \in$

**Algorithm 3.** Wait&Sticking$[b, c, e, \mathbf{o}[1..b-1], k](t, \mathbf{o}[b..t-1])$

**parameter**: $b, c, e \in \mathbb{N}$: beginning, switch, and end times, $1 \le b \le c \le e$
$\mathbf{o}[1..b-1]$: players observations from time 1 to time $b-1$
$k$: positive integer at most the number of arms with no loss.
**input** : $t \in \mathbb{N}$: $b \le t \le e$
$\mathbf{o}[b..t-1]$: players observations from time $b$ to time $t-1$
**output** : $\boldsymbol{\ell}_t \in \{0,1\}^K$: loss vector at time $t$

$S \leftarrow$ the set of arms with no loss by time $b-1$
**if** $t < c$ **then**
  **return** $\mathbb{1}_{[K]\setminus S}$
**else**
  **if** $t = c$ **then**
    $$s_* \leftarrow \arg\max_{s \subseteq S, |s|=k} P\left\{ \sum_{t'=c}^{e} \ell_{t', I_{t'}} \ge 1 \,\middle|\, \begin{array}{l} \mathbf{O}[1..c-1] = \mathbf{o}[1..c-1], \\ \boldsymbol{\ell}_c = \cdots = \boldsymbol{\ell}_e = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\}$$
  **end**
  **return** $\mathbb{1}_{([K]\setminus S)\cup s_*}$
**end**

$[t_{i-1}..t_i]$ to switch the loss vector using BestSwitchingTime$(t_{i-1}, t_i, \mathbf{o}[1..t_{i-1}-1], k_i)$ (Algorithm 2). At this time $t = c$, the adversary calculates the best subset $s \subseteq S$ to add to $[K] \setminus S$ and changes the loss vector $\boldsymbol{\ell}_t$ to $\mathbb{1}_{([K]\setminus S)\cup s}$ using Wait&Sticking$[t_{i-1}, c, t_i, \mathbf{o}[1..t_{i-1}-1], k](t, \mathbf{o}[b..t-1])$ (Algorithm 3).

In RepeatW&S$[K, T]$, $[T]$ and $K-1$ is divided by the solution of Problem 2 shown below. We introduce notation $F(T', K', k)$ for simple description of the problem: for any three integers $T' \ge 1$, $K' \ge 2$ and $k \ge 0$, define function $F(T', K', k)$ as

$$F(T', K', k) = \frac{T' + \binom{K'-1}{k-1} - 1}{T' + \binom{K'}{k} - 1}.$$

Consider the following problem.

*Problem 2.* Given two integers $T \ge 1$ and $K \ge 2$, find two non-negative integer sequences $t_0, \ldots, t_m$ and $k_1, \ldots, k_m$ that maximize

$$\sum_{i=1}^{m} F(t_i - t_{i-1}, K - \sum_{j=1}^{i-1} k_j, k_i)$$

subject to

$$1 \le m \le K-1, \tag{1}$$
$$1 = t_0 < t_1 < \cdots < t_m = T+1 \text{ and} \tag{2}$$
$$k_1 + \cdots + k_m = K-1. \tag{3}$$

*Example 3.* Assume that $K = 2$. Then $m = 1$, $t_0 = 1$, $t_1 = T + 1$, $k_1 = 1$ is the solution of Problem 2. The best time $c^*$ to switch is 2 ($p_1 = 1/2, p_2 = \cdots = p_T = 1$) for the randomized consistent *conservative* player and 1 ($p_c = 1-(1/2)^{T+1-c}$) for the randomized consistent player.

## 4 Lower Loss Bound

**Lemma 4.** *Let $b$ and $T'$ be arbitrary positive integers and let $k$ be an arbitrary non-negative integer. Let $c^*$ be the returned value from BestSwitchingTime$(b, b+T'-1, \mathbf{o}[1..b-1], k)$. Then, for any player algorithm, the following holds with respect to the loss vectors generated by Wait&Sticking$[b, c^*, b+T'-1, \mathbf{o}[1..b-1]](t, \mathbf{o}[b..t-1])$:*

$$\mathbb{E}_{I[b..b+T'-1]}\left[\sum_{t=b}^{b+T'-1} \ell_{t,I_t}\right] \geq F(T', |S|, k),$$

*where $S$ is the set of arms with no loss by time $b-1$, that is, $S = \{i \in [K] : \sum_{t=1}^{b-1} \ell_{t,i} = 0\}$.*

*Remark 5.* In the case with $K = 2$, Lemma 4 implies

$$\mathbb{E}_{I[1..T]}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] \geq F(T, 2, 1) = \frac{T}{T+1}. \tag{4}$$

Equation (4) trivially holds for $T = 1$ because

$$\max_{i\in\{1,2\}} P\left\{\ell_{1,I_1} \geq 1 | \boldsymbol{\ell}_1 = \mathbb{1}_{\{i\}}\right\} = \max_{i\in\{1,2\}} P\left\{I_1 = i\right\} \geq 1/2.$$

Let $p_1 = P\{I_1 = 1\}$, $p_{11} = P\{I_2 = 1 | o_1 = (1, 0)\}$ and $p_{21} = P\{I_2 = 1 | o_1 = (2, 0)\}$. Then, the maximum of the three probabilities

$$P\{\ell_{1,I_1} + \ell_{2,I_2} \geq 1 | \boldsymbol{\ell}_1 = \boldsymbol{\ell}_2 = \mathbb{1}_{\{1\}}\} = P\{I_1 = 1 \text{ or } I_2 = 1\} = p_1 + (1 - p_1)p_{21},$$
$$P\{\ell_{1,I_1} + \ell_{2,I_2} \geq 1 | \boldsymbol{\ell}_1 = \boldsymbol{\ell}_2 = \mathbb{1}_{\{2\}}\} = (1 - p_1) + p_1(1 - p_{11})$$

and

$$\begin{aligned}
&\mathbb{E}_{I_1}\left[\max_{i\in\{1,2\}} P\{\ell_{2,I_2} \geq 1 | I_1, \boldsymbol{\ell}_1 = \mathbf{0}, \boldsymbol{\ell}_2 = \mathbb{1}_{\{i\}}\}\right] \\
&= \mathbb{E}_{I_1}\left[\max_{i\in\{1,2\}} P\{I_2 = i | I_1, \boldsymbol{\ell}_1 = \mathbf{0}\}\right] \\
&= p_1 \max\{p_{11}, 1 - p_{11}\} + (1 - p_1) \max\{p_{21}, 1 - p_{21}\}
\end{aligned}$$

is at least 2/3 because the sum of them is at least 2. The above probabilities are lower bounds of $\mathbb{E}[\ell_{1,I_1} + \ell_{2,I_2}]$ for the cases with $\boldsymbol{\ell}_1 = \boldsymbol{\ell}_2 = \mathbb{1}_{\{1\}}$, $\boldsymbol{\ell}_1 = \boldsymbol{\ell}_2 = \mathbb{1}_{\{2\}}$ and $\boldsymbol{\ell}_1 = \mathbf{0}$ and $\boldsymbol{\ell}_2 = \mathbb{1}_{\{i^*\}}$, respectively, where $i^* = \arg\max_{i\in\{1,2\}} P\{I_2 = i | I_1\}$. Thus, Eq. (4) also holds for $T = 2$. This idea of the proof can be extended to that of Lemma 4. □

*Proof of Lemma 4*. Let

$$\begin{aligned} & p_{b,s}(\mathbf{o}[1..b-1], T') \\ =& P\left\{ \sum_{t=b}^{b+T'-1} \ell_{t,I_t} \geq 1 \;\middle|\; \begin{array}{l} \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \\ \boldsymbol{\ell}_b = \cdots = \boldsymbol{\ell}_{b+T'-1} = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\} \end{aligned}$$

and for any positive integer $c \in [b, b+T'-1]$, let $p_c(\mathbf{o}[1..b-1], T')$ denote the value $p_c$ that is set at Line 4 of BestSwitchingTime$(b, b+T'-1, \mathbf{o}[1..b-1], k)$. Note that,

$$p_b(\mathbf{o}[1..b-1], T') = \max_{s \subseteq S, |s|=k} p_{b,s}(\mathbf{o}[1..b-1], T')$$

holds. Then, with respect to the loss vectors generated by Wait&Sticking $[b, c^*, b+T'-1, \mathbf{o}[1..b-1]](t, \mathbf{o}[b..t-1])$,

$$p_{c^*}(\mathbf{o}[1..b-1], T') \leq \mathbb{E}_{I[b..b+T'-1]} \left[ \sum_{t=b}^{b+T'-1} \ell_{t,I_t} \right]$$

holds. We prove

$$p_{\max} = \max_{t=b,\ldots,e} p_t(\mathbf{o}[1..b-1], T') \geq F(T', |S|, k),$$

which is implied from the inequality

$$\sum_{s \subseteq S, |s|=k} p_{b,s}(\mathbf{o}[1..b-1], T') + \sum_{c=b+1}^{b+T'-1} p_c(\mathbf{o}[1..b-1], T') \geq T' + \binom{|S|-1}{k-1} - 1 \tag{5}$$

because the average of the term values in the left hand side of the inequality is at least

$$\left( T' + \binom{|S|-1}{k-1} - 1 \right) \Big/ \left( T' + \binom{|S|}{k} - 1 \right)$$

if Eq. (5) holds. We prove Eq. (5) for any positive integer $b$ by mathematical induction on $T'$. When $T' = 1$,

$$\begin{aligned} p_{b,s}(\mathbf{o}[1..b-1], 1) =& P\left\{ \ell_{b,I_b} \geq 1 | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \boldsymbol{\ell}_b = \mathbb{1}_{([K]\setminus S)\cup s} \right\} \\ =& P\left\{ I_b \in ([K] \setminus S) \cup s | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1] \right\} \end{aligned}$$

holds. Thus, Eq. (5) holds for any positive integer $b$ because

$$\begin{aligned} & \sum_{s \subseteq S, |s|=k} p_{b,s}(\mathbf{o}[1..b-1], 1) + \sum_{c=b+1}^{b} p_c(\mathbf{o}[1..b-1], 1) \\ =& \sum_{s \subseteq S, |s|=k} P\left\{ I_b \in ([K] \setminus S) \cup s | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1] \right\} \\ \geq& \binom{|S|-1}{k-1} = \binom{|S|-1}{k-1} + T' - 1. \end{aligned}$$

Here, the above inequality holds because each arm $i \in S$ belongs to at least $\binom{|S|-1}{k-1}$ size-$k$ subsets of $S$. Assume that Eq. (5) holds when $T' = n$ for any positive integer $b$. Then,

$$
\begin{aligned}
&p_{b,s}(\mathbf{o}[1..b-1], n+1) \\
=&P\left\{ \sum_{t=b}^{b+n} \ell_{t,I_t} \ge 1 \,\middle|\, \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \boldsymbol{\ell}_b = \cdots = \boldsymbol{\ell}_{b+n} = \mathbb{1}_{([K]\setminus S)\cup s} \right\} \\
=&P\left\{ \ell_{b,I_b} \ge 1 | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \boldsymbol{\ell}_b = \mathbb{1}_{([K]\setminus S)\cup s} \right\} \\
&+ P\left\{ \ell_{b,I_b} < 1, \sum_{t=b+1}^{b+n} \ell_{t,I_t} \ge 1 \,\middle|\, \begin{array}{l} \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \\ \boldsymbol{\ell}_b = \cdots = \boldsymbol{\ell}_{b+n} = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\} \\
=&P\{I_b \in ([K]\setminus S)\cup s | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1]\} \\
&+ P\left\{ I_b \notin ([K]\setminus S)\cup s, \sum_{t=b+1}^{b+n} \ell_{t,I_t} \ge 1 \,\middle|\, \begin{array}{l} \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \\ \boldsymbol{\ell}_b = \mathbb{1}_{[K]\setminus S}, \\ \boldsymbol{\ell}_{b+1} = \cdots = \boldsymbol{\ell}_{b+n} = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\} \\
=&P\{I_b \in ([K]\setminus S)\cup s | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1]\} \\
&+ P\left\{ \sum_{t=b+1}^{b+n} \ell_{t,I_t} \ge 1 \,\middle|\, \begin{array}{l} \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \boldsymbol{\ell}_b = \mathbb{1}_{[K]\setminus S}, \\ \boldsymbol{\ell}_{b+1} = \cdots = \boldsymbol{\ell}_{b+n} = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\} \\
&- P\left\{ I_b \in ([K]\setminus S)\cup s, \sum_{t=b+1}^{b+n} \ell_{t,I_t} \ge 1 \,\middle|\, \begin{array}{l} \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1], \\ \boldsymbol{\ell}_b = \mathbb{1}_{[K]\setminus S}, \\ \boldsymbol{\ell}_{b+1} = \cdots = \boldsymbol{\ell}_{b+n} = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\} \\
=&P\{I_b \in ([K]\setminus S)\cup s | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1]\} \\
&+ \mathbb{E}_{I_b}[p_{b+1,s}(\mathbf{o}[1..b-1], (I_b, \mathbb{1}_{[K]\setminus S, I_b}), n)] \\
&- \sum_{i \in [K]\setminus S \cup s} \begin{array}{l} P\{I_b = i | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1]\} \\ \times p_{b+1,s}(\mathbf{o}[1..b-1], (i, \mathbb{1}_{[K]\setminus S, i}), n) \end{array}
\end{aligned}
$$

and

$$p_c(\mathbf{o}[1..b-1], n+1) = \mathbb{E}_{I_b}\left[p_c(\mathbf{o}[1..b-1], (I_b, \mathbb{1}_{[K]\setminus S, I_b}), n)\right]$$

holds for $c \ge b+1$. Therefore,

$$
\begin{aligned}
&\sum_{s\subseteq S, |s|=k} p_{b,s}(\mathbf{o}[1..b-1], n+1) + \sum_{c=b+1}^{b+n} p_c(\mathbf{o}[1..b-1], n+1) \\
=& \sum_{s\subseteq S, |s|=k} \Big( P\{I_b \in ([K]\setminus S)\cup s | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1]\} \\
&\qquad + \mathbb{E}_{I_b}\left[p_{b+1,s}(\mathbf{o}[1..b-1], (I_b, \mathbb{1}_{[K]\setminus S, I_b}), n)\right] \Big) \\
&- \sum_{i \in [K]\setminus S \cup s} \begin{array}{l} P\{I_b = i | \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1]\} \\ \times p_{b+1,s}(\mathbf{o}[1..b-1], (i, \mathbb{1}_{[K]\setminus S, i}), n) \end{array}
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{c=b+1}^{b+n} \mathbb{E}_{I_b} \left[ p_c(\mathbf{o}[1..b-1], (I_b, \mathbb{1}_{[K]\setminus S, I_b}), n) \right] \\
\geq &\mathbb{E}_{I_b} \Big\{ \sum_{s \subseteq S, |s|=k} p_{b+1,s}(\mathbf{o}[1..b-1], (I_b, \mathbb{1}_{[K]\setminus S, I_b}), n) \\
&\qquad + \sum_{c=b+2}^{b+n} p_c(\mathbf{o}[1..b-1], (I_b, \mathbb{1}_{[K]\setminus S, I_b}), n) \Big\} \\
&+ \sum_{i \in S} P\{ \; I_b = i | \, \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1] \} \\
&\qquad \times \Big\{ \sum_{i \in s \subseteq S, |s|=k} (1 - p_{b+1,s}(\mathbf{o}[1..b-1], (i, \mathbb{1}_{[K]\setminus S, i}), n)) \\
&\qquad\qquad + \max_{s \subseteq S, |s|=k} p_{b+1,s}(\mathbf{o}[1..b-1], (i, \mathbb{1}_{[K]\setminus S, i}), n) \Big\} \\
&+ \sum_{i \in [K]\setminus S} P\{ \; I_b = i | \, \mathbf{O}[1..b-1] = \mathbf{o}[1..b-1] \} \\
&\qquad \times \Big\{ \sum_{s \subseteq S, |s|=k} (1 - p_{b+1,s}(\mathbf{o}[1..b-1], (i, \mathbb{1}_{[K]\setminus S, i}), n)) \\
&\qquad\qquad + \max_{s \subseteq S, |s|=k} p_{b+1,s}(\mathbf{o}[1..b-1], (i, \mathbb{1}_{[K]\setminus S, i}), n) \Big\} \\
\geq &\mathbb{E}_{I_b} \left[ \binom{|S|-1}{k-1} + n - 1 \right] + 1 = \binom{|S|-1}{k-1} + (n+1) - 1
\end{aligned}
$$

holds, which means Eq. (5) holds for any positive integer $b$ when $T' = n + 1$. Note that the last inequality holds by the assumption that Eq. (5) holds for any positive integer $b$ when $T' = n$. □

The following theorem is trivial by Lemma 4.

**Theorem 6.** *RepeatW&S[K, T] forces any player algorithm to suffer the expected loss of at least*

$$
\sum_{i=1}^{m} F(t_i - t_{i-1}, K - \sum_{j=1}^{i-1} k_j, k_i)
$$

*for any positive integers $t_0, \ldots, t_m$, $k_1, \ldots, k_m$ that satisfies (1),(2) and (3).*

**Corollary 7.** *RepeatW&S[K, T] forces any player algorithm to suffer the expected loss of at least*

$$
T(1 - K^{-1/T}) - \frac{1 - K^{(T-1)/T}}{K(1 - K^{1/T})} \tag{6}
$$

*for $T \leq K - 2$,*

$$
H_K - H_{h+1} + h - \frac{A^2(h)(B(h) + 4h)}{2B^2(h)} \tag{7}
$$

*for* $T \geq \frac{h(h-1)}{2} + K - 1$ $(h = 1, \ldots, K-1)$*, where* $H_n$ *is the* $n$*th harmonic number for any natural number* $n$ *and*

$$\begin{aligned} A(h) =& 2\sum_{j=1}^{h} \sqrt{j} \\ B(h) =& 2T - 2(K-1) + (h+3)h. \end{aligned}$$

*Proof.* For $T \leq K - 2$, consider positive integers $t_0, \ldots, t_m$ that satisfy

$$m = T \text{ and } t_1 = 2, t_2 = 3, ..., t_{m-1} = T, t_m = T + 1$$

Then, in this case,

$$\begin{aligned} \sum_{i=1}^{m} F(t_i - t_{i-1}, K - \sum_{j=1}^{i-1} k_j, k_i) =& \sum_{i=1}^{T} F(1, K - \sum_{j=1}^{i-1} k_j, k_i) \\ =& \sum_{i=1}^{T} \frac{\binom{K - \sum_{j=1}^{i-1} k_j - 1}{k_i - 1}}{\binom{K - \sum_{j=1}^{i-1} k_j}{k_i}} \\ =& \sum_{i=1}^{T} \frac{k_i}{K - \sum_{j=1}^{i-1} k_j} \\ =& T - \sum_{i=1}^{T} \frac{K - \sum_{j=1}^{i} k_j}{K - \sum_{j=1}^{i-1} k_j} \end{aligned}$$

holds. By the inequality of arithmetic and geometric means, we have

$$T - \sum_{i=1}^{T} \frac{K - \sum_{j=1}^{i} k_j}{K - \sum_{j=1}^{i-1} k_j} \leq T - TK^{-1/T}$$

with equality if and only if

$$\frac{K - \sum_{j=1}^{i} k_j}{K - \sum_{j=1}^{i-1} k_j} = K^{-1/T} \tag{8}$$

for all $i = 1, \ldots, T$. Unfortunately, $k_1, \ldots, k_T$ that satisfy (8) are not integers. As an approximate solution, use $k_1, \ldots, k_T$ that satisfy

$$K - \sum_{j=1}^{i} k_j = \lceil K^{(T-i)/T} \rceil,$$

then

$$\begin{aligned} T - \sum_{i=1}^{T} \frac{K - \sum_{j=1}^{i} k_j}{K - \sum_{j=1}^{i-1} k_j} &= T - \sum_{i=1}^{T} \frac{\lceil K^{-(T-i)/T} \rceil}{\lceil K^{-(T-i+1)/T} \rceil} \\ &\geq T - \sum_{i=1}^{T-1} \frac{K^{(T-i)/T} + 1}{K^{(T-i+1)/T}} + \frac{1}{K^{1/T}} \\ &= T(1 - K^{-1/T}) - \frac{1 - K^{(T-1)/T}}{K(1 - K^{1/T})}. \end{aligned}$$

Let $h$ be an arbitrary integer in $[1, K-1]$. Consider the case with $T \geq \frac{h(h-1)}{2} + K - 1$. Let $m = K - 1$ and let

$$t_i = \begin{cases} t_{i-1} + 1 & (i = 1, \dots, K - h - 1) \\ t_{i-1} + T_i & (i = K - h, \dots, K - 1) \end{cases}$$

and $k_1 = \cdots = k_{K-1} = 1$, where $T_i$ is non-negative integer with

$$\sum_{i=1}^{h} T_{K-i} = T - (K-1). \tag{9}$$

Then,

$$\begin{aligned} &\sum_{i=1}^{m} F(t_i - t_{i-1}, K - \sum_{j=1}^{i-1} k_j, k_i) \\ =& \sum_{i=1}^{K-h-1} F(1, K-i+1, 1) + \sum_{i=K-h}^{K-1} F(T_i + 1, K - i + 1, 1) \\ =& \sum_{i=1}^{K-h-1} \frac{1}{K-i+1} + \sum_{i=K-h}^{K-1} \frac{1+T_i}{K-i+1+T_i} \\ =& H_K - H_{h+1} + h - \sum_{i=K-h}^{K-1} \frac{K-i}{K-i+1+T_i} \\ =& H_K - H_{h+1} + h - \sum_{i=1}^{h} \frac{i}{i+1+T_{K-i}} \end{aligned} \tag{10}$$

holds. Let

$$f(T_{K-h}, \dots, T_{K-1}) = \sum_{i=1}^{h} \frac{i}{i+1+T_{K-i}}.$$

By solving the problem of maximizing $f(T_{K-h}, \dots, T_{K-1})$ subject to Constraint (9) using the method of Lagrange multipliers, we obtain

$$T_{K-i} = \frac{B(h)\sqrt{i}}{A(h)} - (i+1) \text{ for } i = 1, \dots, h. \tag{11}$$

All the $T_{K-i}$ are non-negative because

$$\begin{aligned}\frac{B(h)\sqrt{i}}{A(h)}-(i+1) =&\frac{2T-2(K-1)+h(h+3)}{2\sum_{j=1}^{h}\sqrt{j}}\sqrt{i}-(i+1)\\ \geq&\frac{2h(h+1)}{2\sum_{j=1}^{h}\sqrt{j}}\sqrt{i}-(i+1)\\ \geq&\frac{2h(h+1)}{h\sqrt{2(h+1)}}\sqrt{i}-(i+1)\\ =&\sqrt{2(h+1)i}-(i+1)\geq 0\end{aligned}$$

holds for $i=1,\ldots,h$. Here, the first inequality holds because $T\geq\frac{h(h-1)}{2}+K-1$ and the second inequality holds by inequality $\sum_{j=1}^{h}\sqrt{j}\leq h\sqrt{\frac{h+1}{2}}$. Due to integer constraint, instead of $T_{K-i}$ defined by Eq. (11), we use $T_{K-i}$ defined as follows:

$$i+1+T_{K-i}=\left\lfloor\sum_{j=1}^{i}\frac{B(h)\sqrt{j}}{A(h)}\right\rfloor-\left\lfloor\sum_{j=1}^{i-1}\frac{B(h)\sqrt{j}}{A(h)}\right\rfloor.$$

Then,

$$\begin{aligned}\sum_{i=1}^{h}\frac{i}{i+1+T_{K-i}} <&\sum_{i=1}^{h}\frac{i}{\frac{B(h)}{A(h)}\sqrt{i}-1}\\ =&A(h)\sum_{i=1}^{h}\frac{i}{B(h)\sqrt{i}-A(h)}\\ =&\frac{A^2(h)}{2B(h)}+\frac{A^2(h)}{B^2(h)}h+\frac{A^3(h)}{B^2(h)}\sum_{i=1}^{h}\frac{1}{B(h)\sqrt{i}-A(h)}\\ \leq&\frac{A^2(h)}{2B(h)}+\frac{A^2(h)}{B^2(h)}h+\frac{A^2(h)}{B^2(h)}h\\ =&\frac{A^2(h)(B(h)+4h)}{2B^2(h)}\end{aligned}\tag{12}$$

holds. Here, the first inequality uses

$$\left\lfloor\sum_{j=1}^{i}\frac{B(h)\sqrt{j}}{A(h)}\right\rfloor-\left\lfloor\sum_{j=1}^{i-1}\frac{B(h)\sqrt{j}}{A(h)}\right\rfloor>\frac{B(h)\sqrt{i}}{A(h)}-1$$

and the second inequality uses the fact that

$$B(h)\sqrt{i}-A(h)\geq B(h)-A(h)\geq A(h),$$

which can be implied from inequalities

$$A(h)\leq h\sqrt{2(h+1)}$$

and

$$B(h) = 2\left\{T - (K-1) - \frac{h(h-1)}{2}\right\} + 2h(h+1) \leq 2h(h+1).$$

By Eqs. (10) and (12), Bound (7) holds in this case. □

**Corollary 8.** *RepeatW&S[K,T] forces any player algorithm to suffer expected loss at least*

$$K - 1 - \frac{K(K-1)^2(2T + (K-1)(K+4))}{(2T + K(K-1))^2}$$

*for* $T \geq K(K-1)/2$.

*Proof.* This corollary can be derived from Bound (7) of Corollary 7 with $h = K - 1$ and the fact that

$$A^2(K-1) = 4\left(\sum_{j=1}^{K-1} \sqrt{j}\right)^2 \leq 4(K-1)\sum_{j=1}^{K-1} j = 2K(K-1)^2.$$

□

## 5 Concluding Remark

In this paper, we focus on the adversary's strategy. Any consistent player that avoids arms with previously observed loss incurs total loss at most $K - 1$. This leaves a small $O(1/T)$ gap between the loss forced by our adversary and this trivial loss bound for consistent players. Finding minimax strategies to close this gap remains an open problem.

**Acknowledgment.** We would like to thank anonymous reviewers for helpful comments. This work was partially supported by JSPS KAKENHI Grant Number 25280079.

## References

1. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.E.: The nonstochastic multi-armed bandit problem. SIAM J. Comput. **32**(1), 48–77 (2003)
2. Bubeck, S., Cesa-Bianchi, N.: Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Found. Trends Mach. Learn. **5**(1), 1–122 (2012)
3. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press, Cambridge (2006)