Noise Free Multi-Armed Bandit Game

Atsuyoshi Nakamura
(Hokkaido University)

David P. Helmbold
(UCSC)

Manfred K. Warmuth
(UCSC)

Noise Free Multi-Armed Bandit Game

for $t = 1, 2, \ldots, T$ **do**

1. The adversary picks a loss $\ell_{t,i} \in [0, 1]$ for each arm $i \in \{1, \ldots, K\}$.
   (The player does not know $\ell_{t,i}$ for any arm i.)
2. The player chooses arm I_t.
3. Loss ℓ_{t,I_t} only is revield and the player suffers ℓ_{t,I_t}.

Player's obj.: minimizing his/her expected cumulative loss.
Adversary's obj.: maximizing player's expected cumulative loss.

Noise free setting: There is an arm which never suffers any loss.

[Assumptions]

- The player can use a randomized strategy.
- The adversary has infinite computation power.
- The adversary can behave adaptively: the adversary's decision can depend on both the player's past choices and the adversary's past decisions.

Full Information Case

[Assumption]	The losses $\ell_{t,i}$ of all the arms are revealed at the end of time t.
[minimax expected loss]	$\sum_{i=2}^{K} \frac{1}{i} = \Theta(\ln K)$
[Adversary's strategy]	Set the loss of an arm with the highest probability of being chosen to 1 and others to 0.
[Player's strategy]	Choose one of the arms with no loss so far randomly with equal probability.

- Once $\ell_{i,t}$ is set to non-zero and $I_t \neq i$, then the player never chooses i as I_t at time $s > t$.
  (In the bandit case, the player may choose it.)

Modification of Noisy-Case Adversary's Strategy

Adversary's strategy Select a best arm i first according to the uniform distribution over $\{1, ..., K\}$, then generate each component $\ell_{t,j}$ of each loss vector $\boldsymbol{\ell}_t$ independently according to Bernoulli distribution with parameter 0 for $j = i$ or parameter ϵ for $j \neq i$.

Theorem (Auer et al. 2003)

For the above adversary' strategy, any player algorithm suffers at least

$$\left(1 - \sqrt{\frac{1}{2}\alpha\left(\frac{K}{T}\right)}\right) K - 1$$

loss for $T > \frac{4}{3}K$ *in expectation, where* $\alpha(x) = -\frac{1}{x}\ln(1-x)$.

- $\alpha(x)$ is increasing function with $\lim_{x \to +0} \alpha(x) = 1$.
- The above bound $\to (1 - 1/\sqrt{2})K - 1$ when $T \to \infty$.
- There is a player algorithm that suffers at most $(K-1)/2$ loss for this adversary.

Main Result

We show an adversary strategy of nose-free multi-armed bandit problem that forces any player algorithm to suffer

$$K - 1 - O(1/T)$$

loss in expectation for K-arm and T-round case.

Function F

For any three integers $T' \geq 1$, $K' \geq 2$ and $K' \geq k \geq 1$,

$$F(T', K', k) \stackrel{\text{def}}{=} \left(T' + \binom{K'-1}{k-1} - 1 \right) \Big/ \left(T' + \binom{K'}{k} - 1 \right).$$

(Example)

$$F(T, 2, 1) = \frac{T}{T+1}$$

Theorem of Loss Lower Bound

Theorem

For noise free K-armed bandit game, there is an adversary's strategy that 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*

$$1 \leq m \leq K - 1, \tag{1}$$

$$1 = t_0 < t_1 < \cdots < t_m = T + 1 \text{ and} \tag{2}$$

$$k_1 + \cdots + k_m = K - 1. \tag{3}$$

When $K = 2$, uniquely determined as

$$m = 1, t_0 = 1, t_1 = T + 1, k_1 = 1.$$

2-Armed Case

Corollary

For noise free 2-armed bandit game, there is an adversary's strategy that forces any player algorithm to suffer the expected loss of at least

$$\frac{T}{T+1} = 1 - \frac{1}{T+1}.$$

General Case

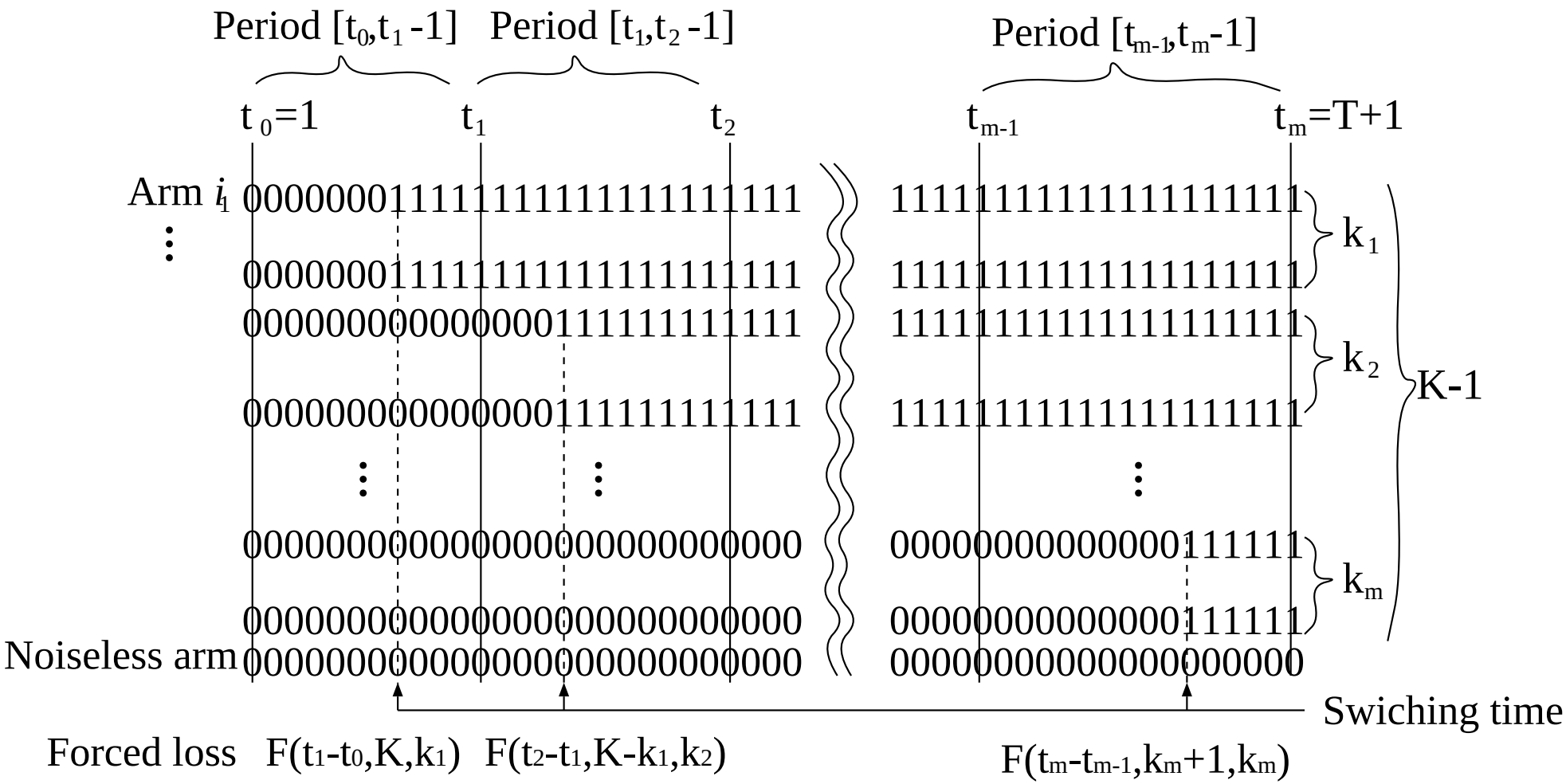
Corollary

For noise free K-armed bandit game, there is an adversary's strategy that forces any player algorithm to suffer the expected loss of at least

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

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

Adversary's Strategy



Partitioning Problem

Problem

For given two integers $T \geq 1$ and $K \geq 2$, 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

$$1 \leq m \leq K - 1, \tag{1}$$

$$1 = t_0 < t_1 < \cdots < t_m = T + 1 \text{ and} \tag{2}$$

$$k_1 + \cdots + k_m = K - 1. \tag{3}$$

Notations

For any natural numbers i, j with $i \leq j$,

$$[i..j] \stackrel{\text{def}}{=} \{i, \ldots, j\}, \quad [j] \stackrel{\text{def}}{=} [1..j]$$

For any sequence $x_1, \ldots, x_n$,

$$\mathbf{x}[b..e] \stackrel{\text{def}}{=} x_b, \ldots, x_e.$$

For losses $\ell_{t,i}$ at time t,

$$\boldsymbol{\ell}_t \stackrel{\text{def}}{=} (\ell_{t,1}, \ldots, \ell_{t,K})$$

i_t: realization of random variable I_t
$O_t \stackrel{\text{def}}{=} (I_t, \ell_{t,I_t})$: player's observation at time t
$o_t \stackrel{\text{def}}{=} (i_t, \ell_{t,I_t})$: realization of O_t
For any set $S \subseteq [K]$,

$$\mathbb{1}_S \stackrel{\text{def}}{=} K\text{-dimensional } \{0,1\}\text{-vector}$$
$$\text{whose } i\text{th component is } 1 \text{ if and only if } i \in S.$$

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 Partitioning Problem

$d \leftarrow 1$

for time $t = 1, \ldots, T$ **do**
  if $t \geq t_{d-1}$ **then**
    $b \leftarrow t_{d-1}$, $e \leftarrow t_d - 1$
    $c \leftarrow$BestSwitchingTime$(b, e, \mathbf{o}[1..b-1], k_d)$
    $d \leftarrow d + 1$
  end if
  $\ell_t \leftarrow$ Wait&Sticking$[b, c, e, \mathbf{o}[1..b-1], k_d](t, \mathbf{o}[b, ..t-1])$
  Observe the player's choice i_t
end for

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

1: $S \leftarrow$ the set of arms with no loss by time $b-1$
2: $p_{\max} = -1$
3: **for** $c = b, \dots, e$ **do**
4: $p_c \leftarrow$

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

5: **if** $p_c > p_{\max}$ **then**
6: $c^* \leftarrow c, \quad p_{\max} \leftarrow p_c$
7: **end if**
8: **end for**
9: **return** c^*

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, $1 \leq b \leq c \leq e$
$\mathbf{o}[1..b-1]$: players observations from time 1 to time $b-1$
k: positive integer at most the number of arms with no loss.

Input: $t \in \mathbb{N}$: $b \leq t \leq e$
$\mathbf{o}[b..t-1]$: 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**
    $$s_* \leftarrow \arg\max_{s \subseteq S, |s|=k} P\left\{ \sum_{t'=c}^{e} \ell_{t', I_{t'}} \geq 1 \,\middle|\, \begin{array}{l} \mathbf{O}[1..c-1] = \mathbf{o}[1..c-1], \\ \boldsymbol{\ell}_c = \cdots = \boldsymbol{\ell}_e = \mathbb{1}_{([K]\setminus S)\cup s} \end{array} \right\}$$
  end if
  return $\mathbb{1}_{([K]\setminus S)\cup s_*}$
end if

Main Lemma

Lemma

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, $S = \{i \in [K] : \sum_{t=1}^{b-1} \ell_{t,i} = 0\}$.

Proof Scketch (1/2)

BestSwitchingTime$(b, e, \mathbf{o}[1..b-1], k)$ returns $c^* = c$ that maximizes

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

Define

$$p_{b,s}(\mathbf{o}[1..b-1], e-b+1) \stackrel{\text{def}}{=} P\left\{\textstyle\sum_{t=b}^{e}\ell_{t,I_t} \geq 1 \middle| \mathbf{O}[1..b-1]=\mathbf{o}[1..b-1], \boldsymbol{\ell}_b=\cdots=\boldsymbol{\ell}_e=\mathbb{1}_{([K]\setminus S)\cup s}\right\}.$$

Then,

$$p_b = \max_{s\subseteq S,|s|=k} p_{b,s}(\mathbf{o}[1..b-1], e-b+1).$$

Proof Scketch (2/2)

Thus,

$$\mathbb{E}_{I[b..e]}\left[\sum_{t=b}^{e}\ell_{t,I_t}\right] \geq p_{c^*}$$
$$= \max\left(\{p_{b,s}(\mathbf{o}[1..b-1], e-b+1) | s \subseteq S, |s| = k\} \cup \{p_c | c = b+1, ..., e\}\right)$$

We show

$$\sum_{s\subseteq S,|s|=k} p_{b,s}(\mathbf{o}[1..b-1], e-b+1) + \sum_{c=b+1}^{e} p_c \geq e-b+1+\binom{|S|-1}{k-1}-1,$$

which implies

$$p_{c^*} \geq \left(e-b+\binom{|S|-1}{k-1}\right) \Big/ \left(e-b+\binom{|S|}{k}\right)$$

Future Work

What is the minimax-strategy of the noise-free multi-armed bandit game?

Minimax Strategy for Noise Free 2-Armed Bandit Game

$i^* \leftarrow ?$
for time $t = 1, \ldots, T$ **do**
  if $i^* = ?$ **then**
    if $t > 1$ and $i_{t-1} = 1$ **then**
      $I_t \leftarrow 1$
    else
      $$i_t \leftarrow \begin{cases} 1 \text{ with prob.} 1/(T-t+2) \\ 2 \text{ with prob.} (T-t+1)/(T-t+2) \end{cases}$$
    end if
  else
    $i_t \leftarrow i^*$
  end if
  ℓ_{t,i_t} is observed.
  if $\ell_{t,i_t} > 0$ **then**
    $$i^* \leftarrow \begin{cases} 1 & (i_t = 2) \\ 2 & (i_t = 1) \end{cases}$$
  end if
end for