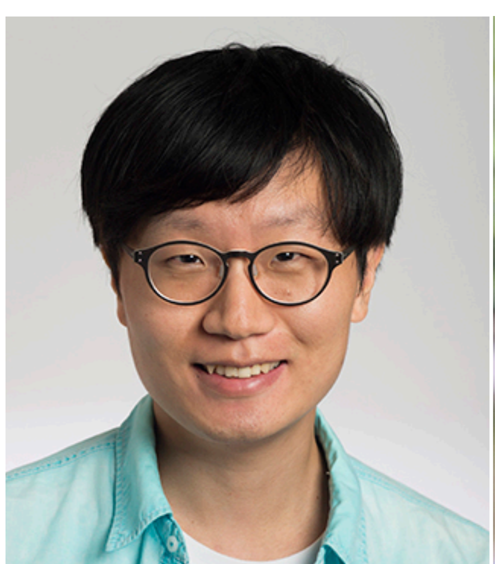
Are sample means in multi-armed bandits positively or negatively biased?

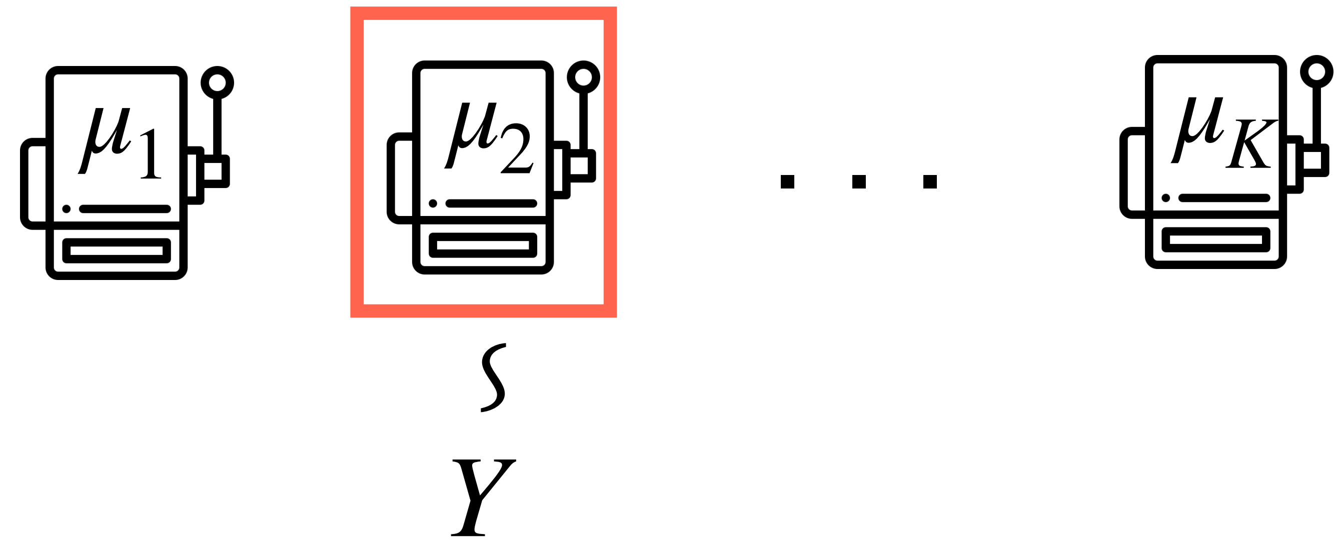
Jaehyeok Shin[1], Aaditya Ramdas[1,2] and Alessandro Rinaldo[1]

Dept. of Statistics and Data Science[1],
Machine Learning Dept.[2], CMU



Poster #12 @ Hall B + C

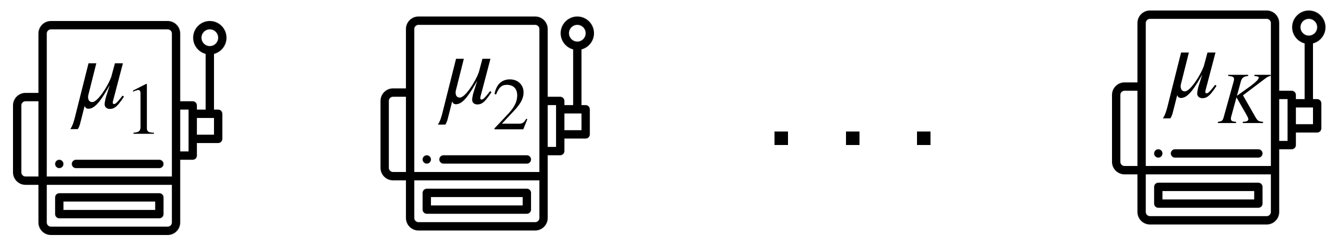
Stochastic multi-armed bandit



"Random reward"

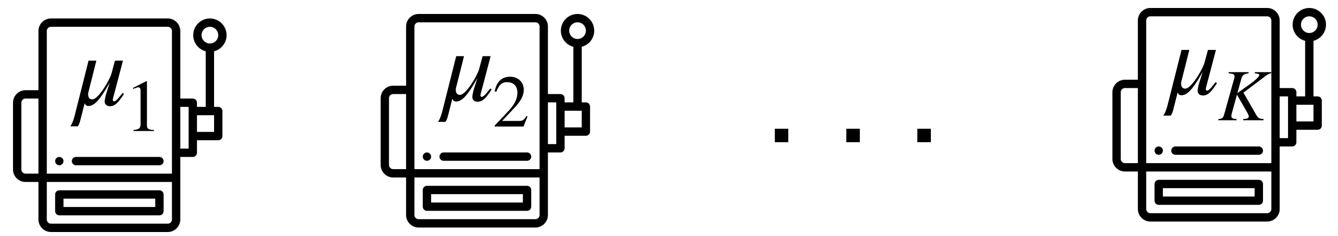
Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time



Adaptive sampling scheme
to maximize rewards / to identify the best arm

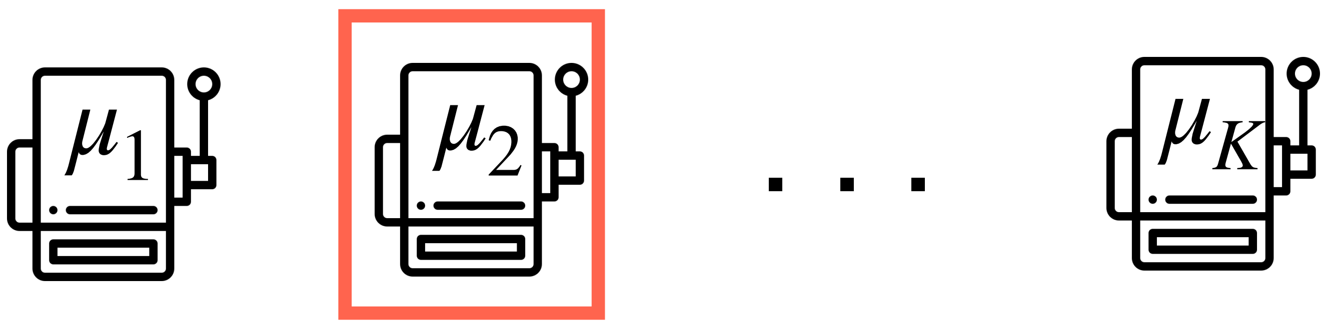
Time



$t = 1$

Adaptive sampling scheme
to maximize rewards / to identify the best arm

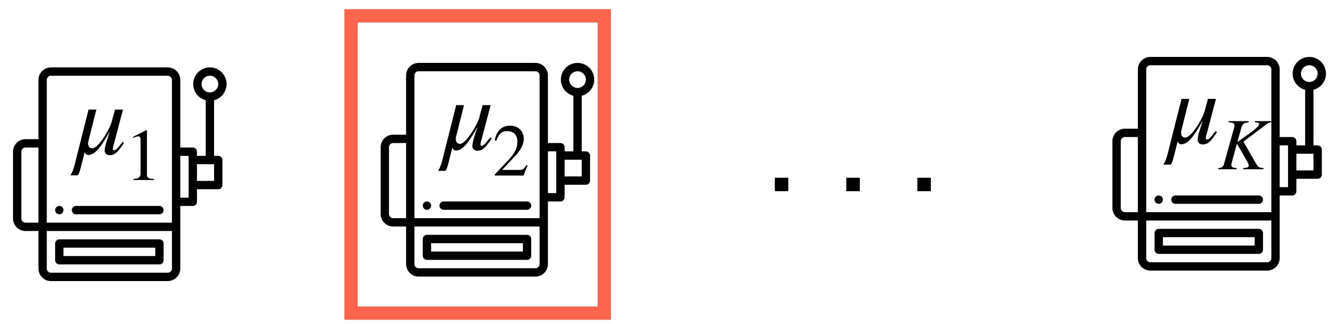
Time



$t = 1$

Adaptive sampling scheme
to maximize rewards / to identify the best arm

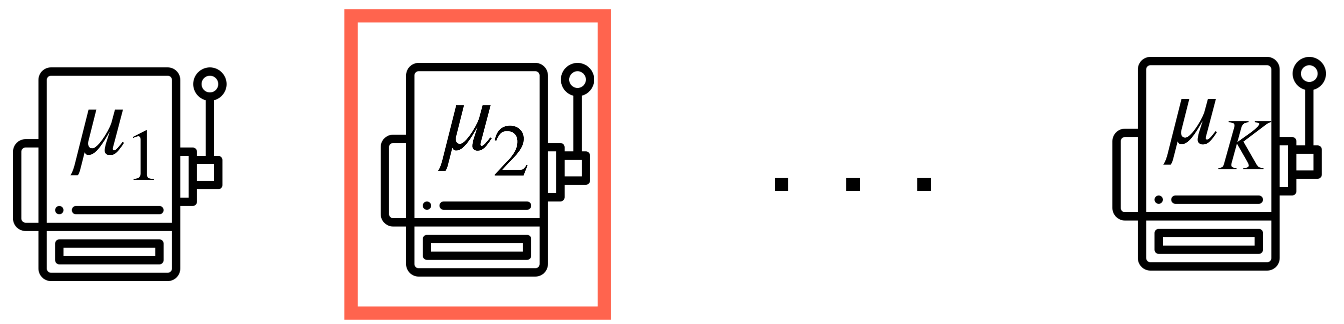
Time



$t = 1$ Y_1

Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time

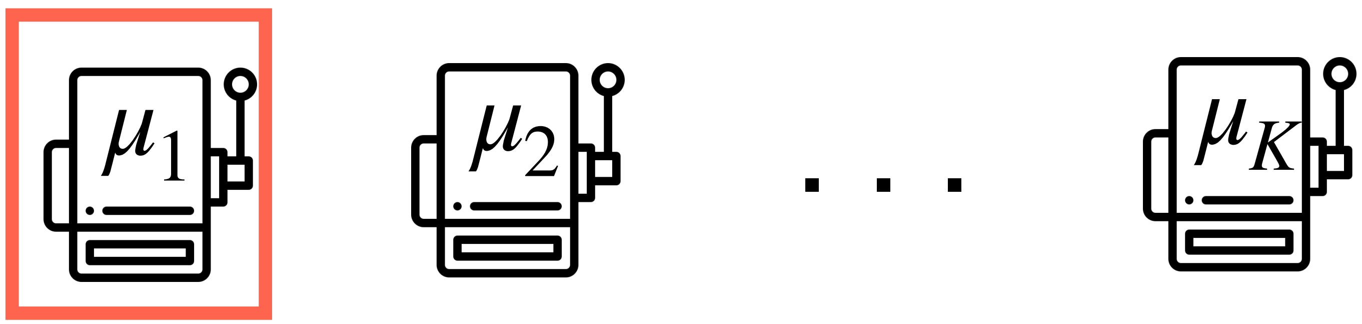


$t = 1$ $\qquad Y_1$

$t = 2$

Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time

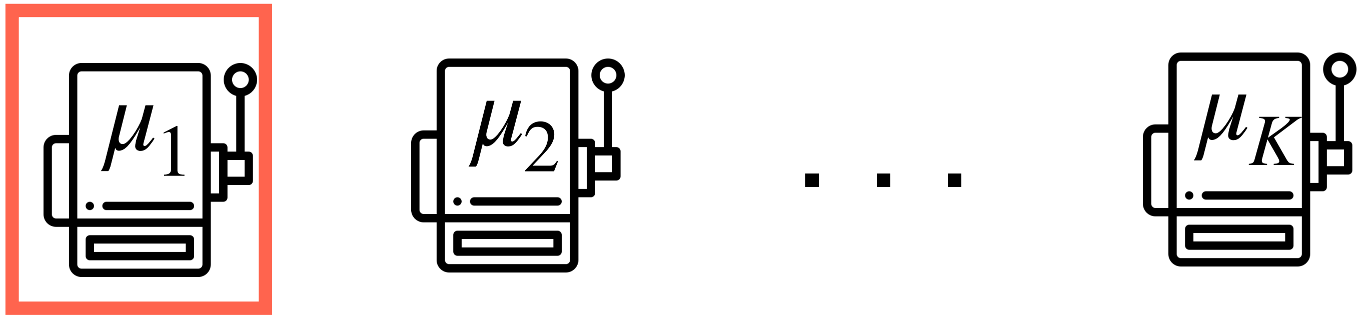


$t = 1$ $\quad$ Y_1

$t = 2$

Adaptive sampling scheme
to maximize rewards / to identify the best arm

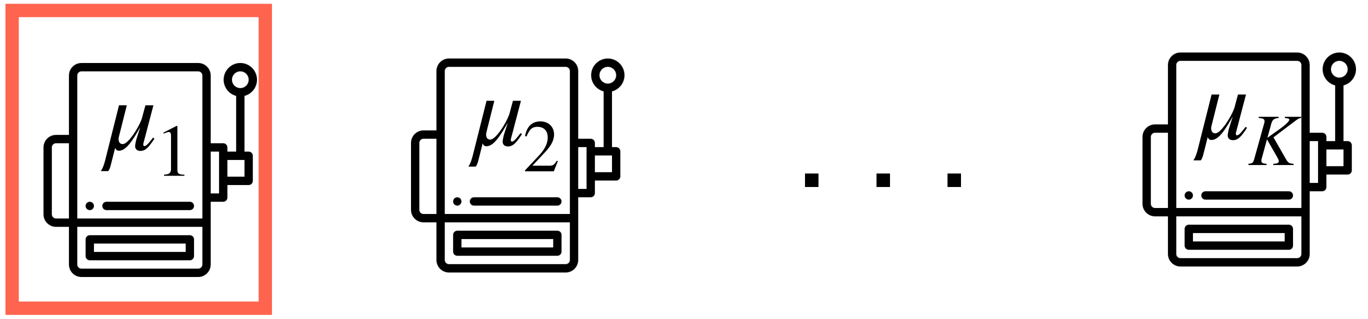
Time



Time	μ_1	μ_2	$\dots$	μ_K
$t = 1$		Y_1		
$t = 2$	Y_2			

Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time



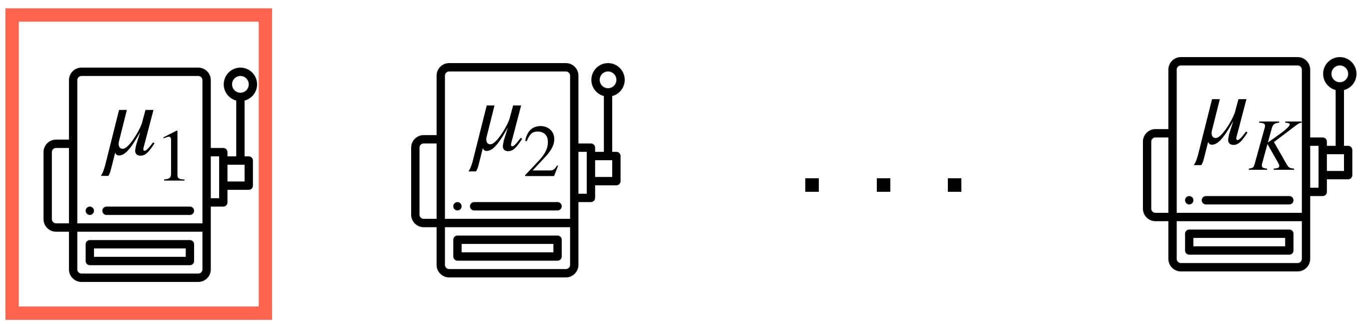
$t = 1$ $\qquad\qquad\qquad\qquad Y_1$

$t = 2$ $\qquad\qquad Y_2$

$\vdots$

Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time



$t = 1$ $\quad Y_1$

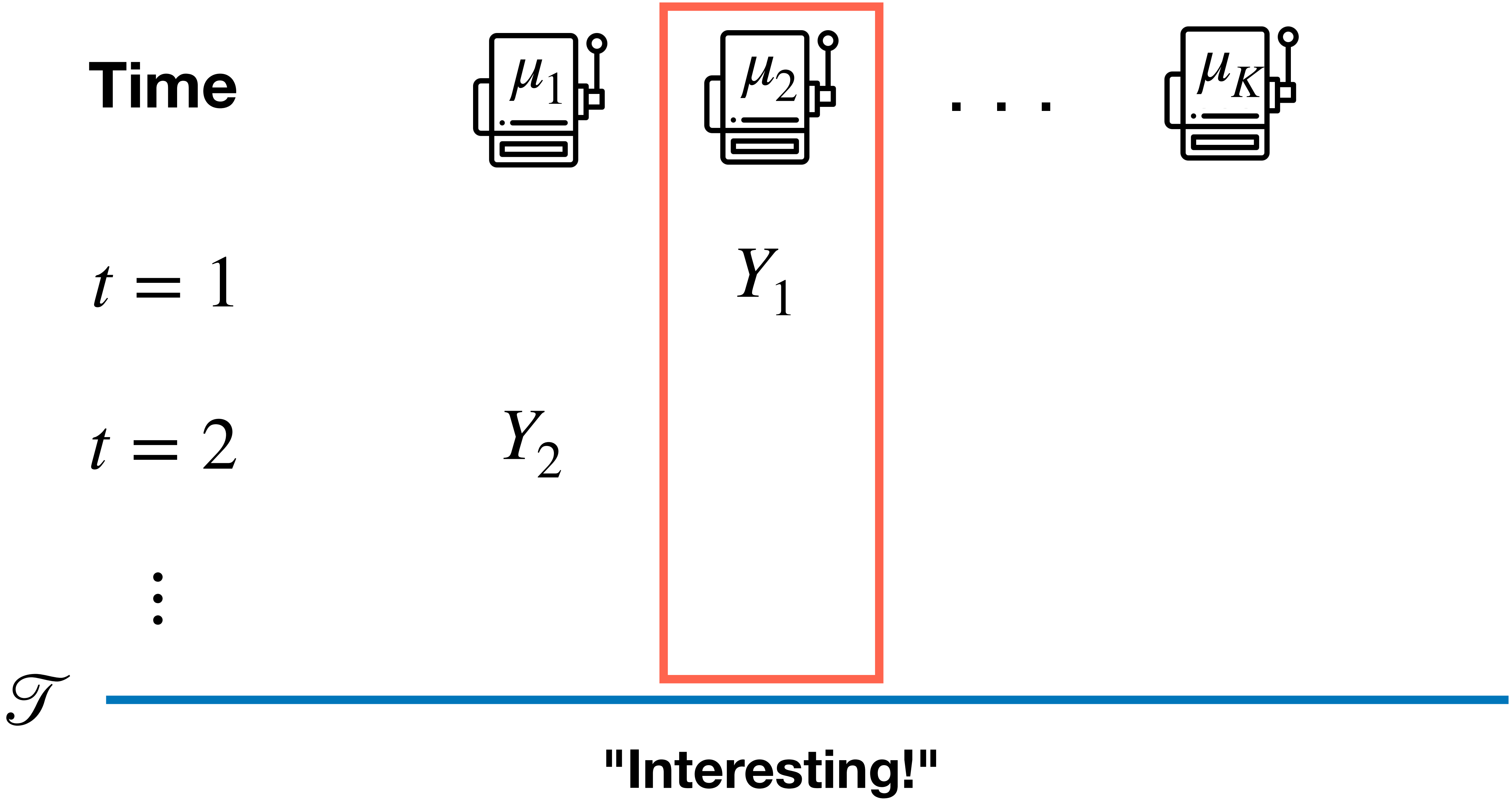
$t = 2$ $\quad Y_2$

$\vdots$

$\mathscr{T}$

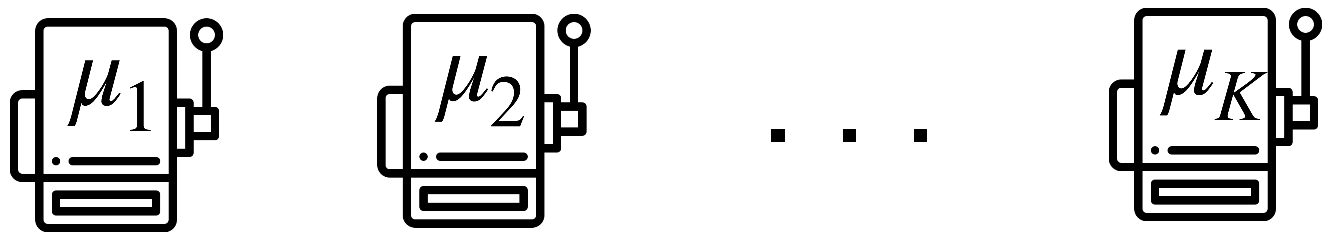
Stopping time

Collected data can be used to identify an interesting arm...



...and data can be used to estimate the mean.

Time



$t = 1$ $\quad Y_1$

$t = 2$ $\quad Y_2$

$\vdots$

$\mathcal{T}$

$\boxed{\widehat{\mu}_\kappa(\mathcal{T})}$ **Sample mean of chosen arm κ**

Q. Bias of sample mean?

$$\mathbb{E}\left[\widehat{\mu}_\kappa(\mathcal{T}) - \mu_\kappa\right] \leq \textbf{ or } \geq 0?$$

Nie et al. 2018

: Sample mean is negatively biased.

$$\mathbb{E}\left[\widehat{\mu}_k(t) - \mu_k\right] \leq 0$$

Nie et al. 2018

: Sample mean is **negatively** biased.

$$\mathbb{E}\left[\widehat{\mu}_k(t) - \mu_k\right] \le 0$$

Fixed Arm **Fixed Time**

Nie et al. 2018

: Sample mean is negatively biased.

$$\mathbb{E}\left[\widehat{\mu}_k(t) - \mu_k\right] \leq 0$$

Fixed Arm

Fixed Time

This work

: Sample mean of chosen arm at stopping time

$$\mathbb{E}\left[\widehat{\mu}_\kappa(\mathcal{T}) - \mu_\kappa\right]$$

Chosen Arm

Stopping Time

<u>This work</u>

: Sample mean of chosen arm at stopping time is ...

$$\mathbb{E}\left[\widehat{\mu}_{\kappa}(\mathcal{T}) - \mu_{\kappa}\right]$$

<u>This work</u>

: Sample mean of chosen arm at stopping time is ...

$$\mathbb{E}\left[\widehat{\mu}_{\kappa}(\mathcal{T}) - \mu_{\kappa}\right]$$

(a) negatively biased under 'optimistic sampling'.

<u>This work</u>

: Sample mean of chosen arm at stopping time is ...

$$\mathbb{E}\left[\widehat{\mu}_{\kappa}(\mathcal{T}) - \mu_{\kappa}\right]$$

(a) negatively biased under 'optimistic sampling'.

(b) positively biased under 'optimistic stopping'.

This work

: Sample mean of chosen arm at stopping time is ...

$$\mathbb{E}\left[\widehat{\mu}_{\kappa}(\mathcal{T}) - \mu_{\kappa}\right]$$

(a) negatively biased under 'optimistic sampling'.

(b) positively biased under 'optimistic stopping'.

(c) positively biased under 'optimistic choosing'.

Monotone effect of a sample

<u>**Theorem**</u> [Informal]

$$\textbf{Sample from arm } k \longmapsto \frac{\mathbf{1}(\kappa = k)}{N_k(\mathcal{T})}$$

Monotone effect of a sample

<u>**Theorem**</u> [Informal]

Sample from arm k $\longmapsto$ $\dfrac{\mathbf{1}(\kappa = k)}{N_k(\mathcal{T})}$ **Positive bias**

Increasing

Monotone effect of a sample

Theorem [Informal]

Sample from arm k $\longmapsto$ $$\frac{\mathbf{1}(\kappa = k)}{N_k(\mathscr{T})}$$

Increasing

Decreasing

Positive bias

Negative bias

Monotone effect of a sample

Theorem [Informal]

Sample from arm k $\longmapsto$ $\dfrac{\mathbf{1}(\kappa = k)}{N_k(\mathscr{T})}$

Increasing — **Positive bias**

Decreasing — **Negative bias**

➡ **Agnostic to algorithm**

Monotone effect of a sample

<u>**Theorem**</u> [Informal]

Sample from arm k $\longmapsto$ $\dfrac{\mathbf{1}(\kappa = k)}{N_k(\mathcal{T})}$

Increasing **Positive bias**

Decreasing **Negative bias**

- **Agnostic to algorithm**
- **Includes Nie et al. 2018 as a special case**

Monotone effect of a sample

Theorem [Informal]

Sample from arm k $\longmapsto$ $\dfrac{\mathbf{1}(\kappa = k)}{N_k(\mathscr{T})}$

Increasing → **Positive bias**

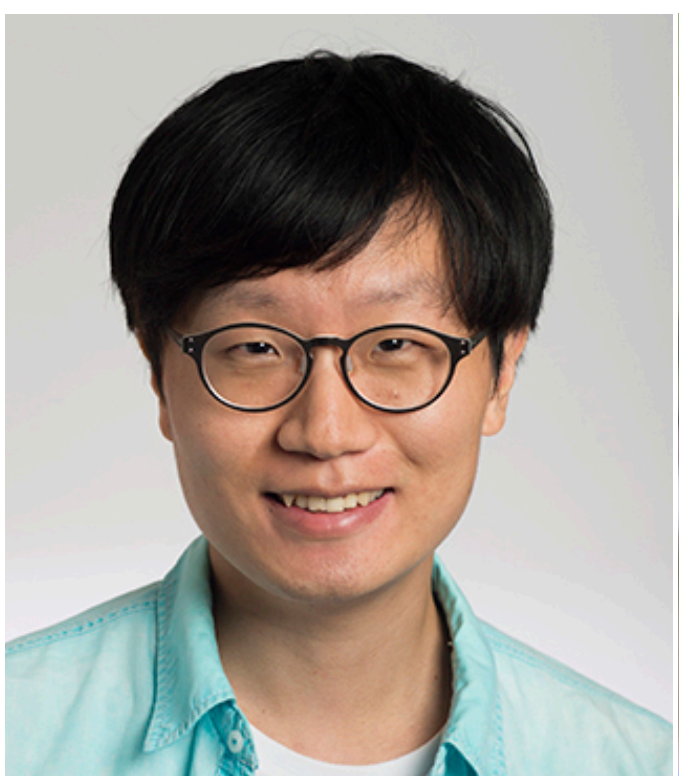
Decreasing → **Negative bias**

- **Agnostic to algorithm**
- **Includes Nie et al. 2018 as a special case**
- **Positive bias under best arm identification, sequential testing**

Poster #12 @ Hall B + C

Are sample means in multi-armed bandits positively or negatively biased?

Jaehyeok Shin, Aaditya Ramdas and Alessandro Rinaldo



Carnegie
Mellon
University