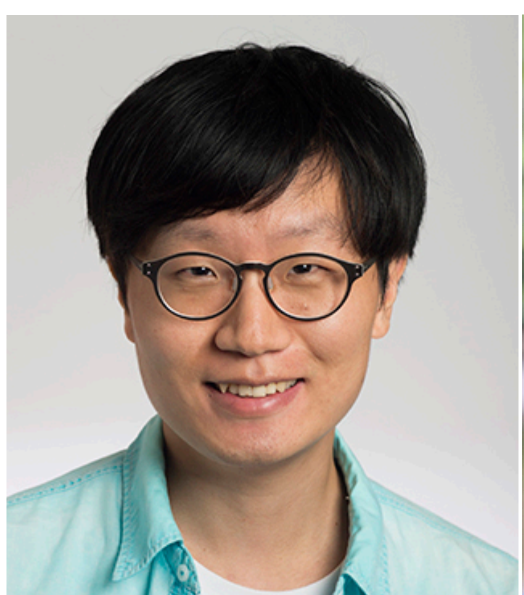
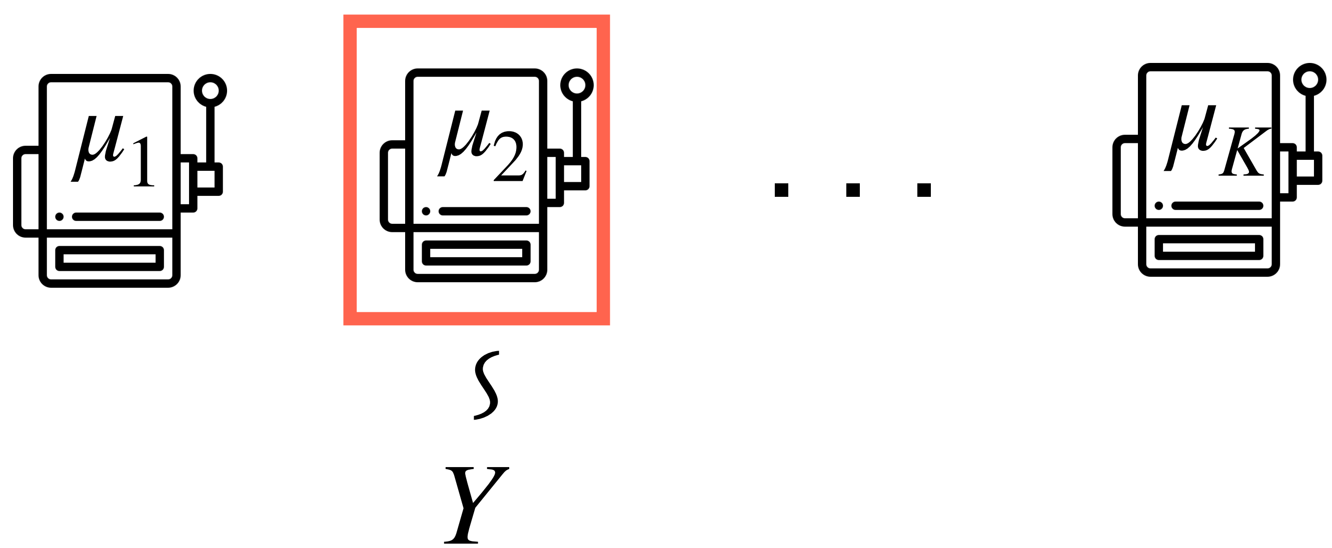
On conditional versus marginal bias in multi-armed bandits

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

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



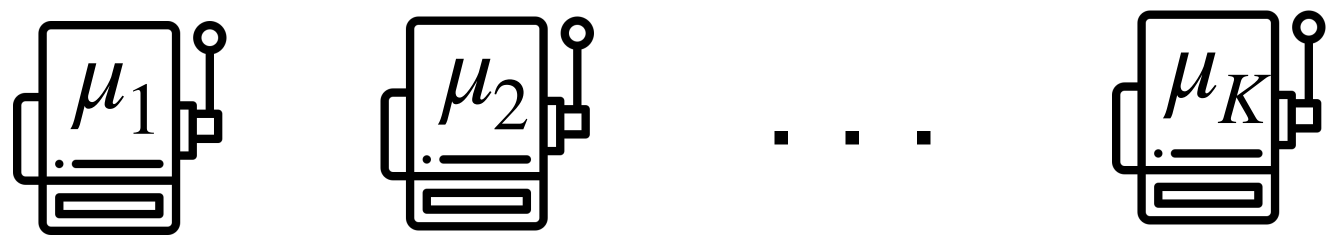
Stochastic Multi-armed bandits (MABs)



"Random reward"

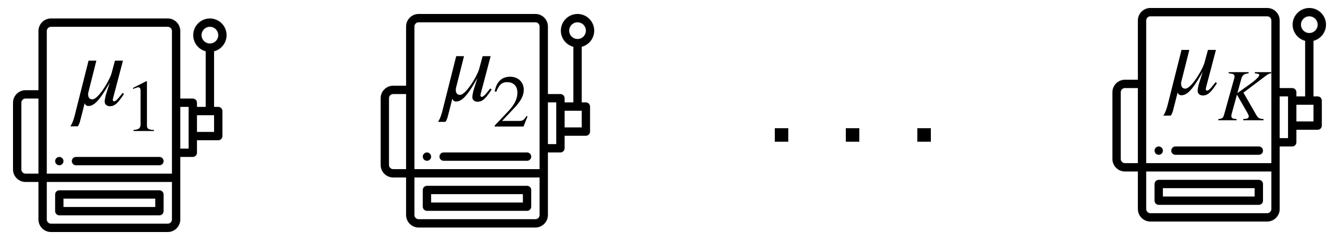
Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time



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

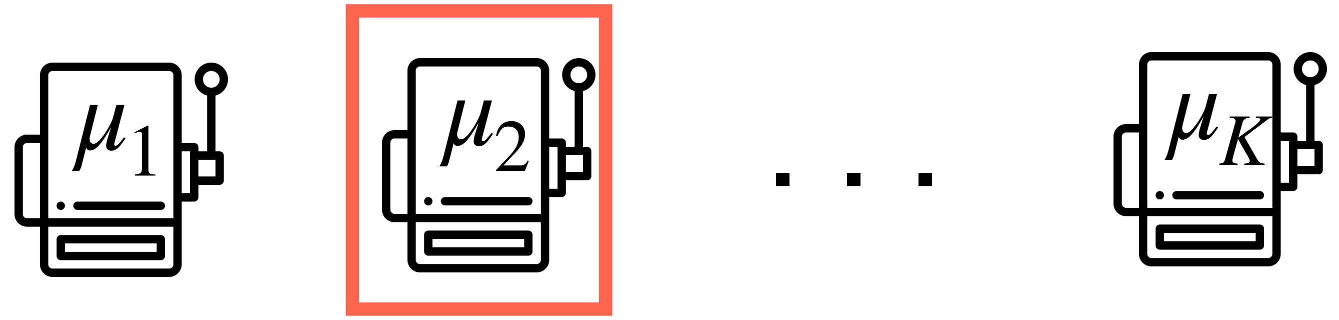
Time



$t = 1$

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

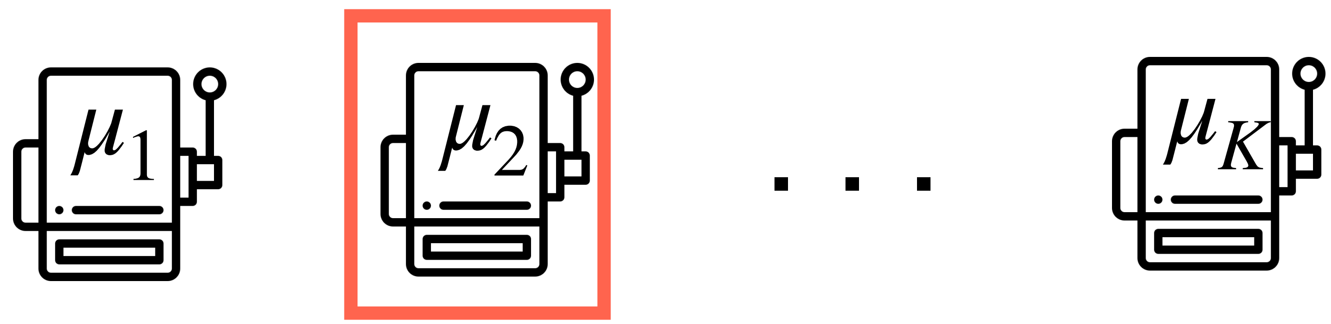
Time



$t = 1$

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

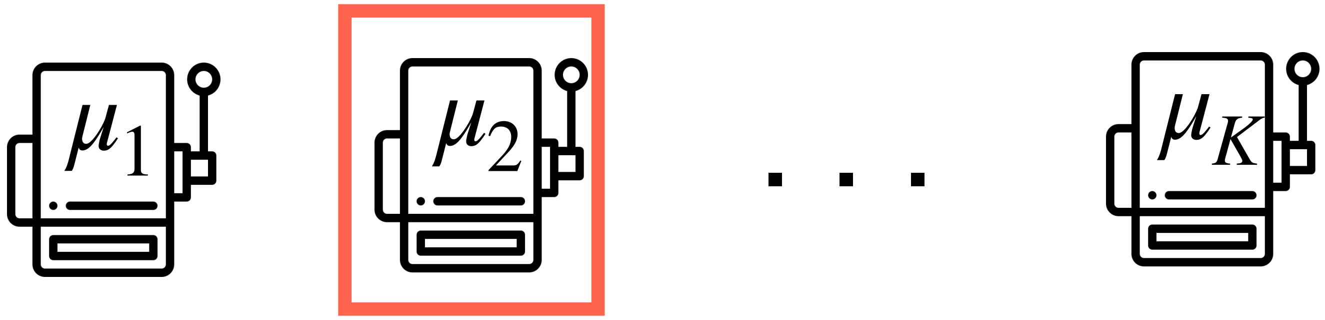
Time



$t = 1$ Y_1

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

Time

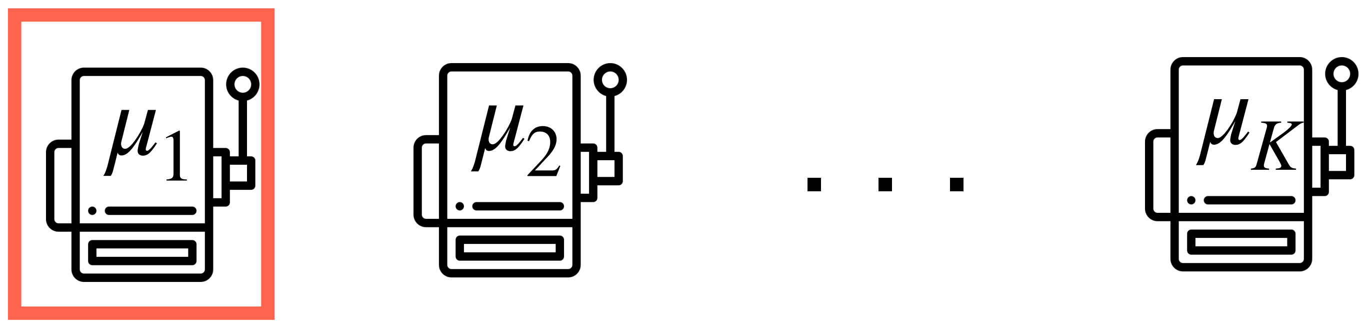


$t = 1$ $\quad Y_1$

$t = 2$

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

Time

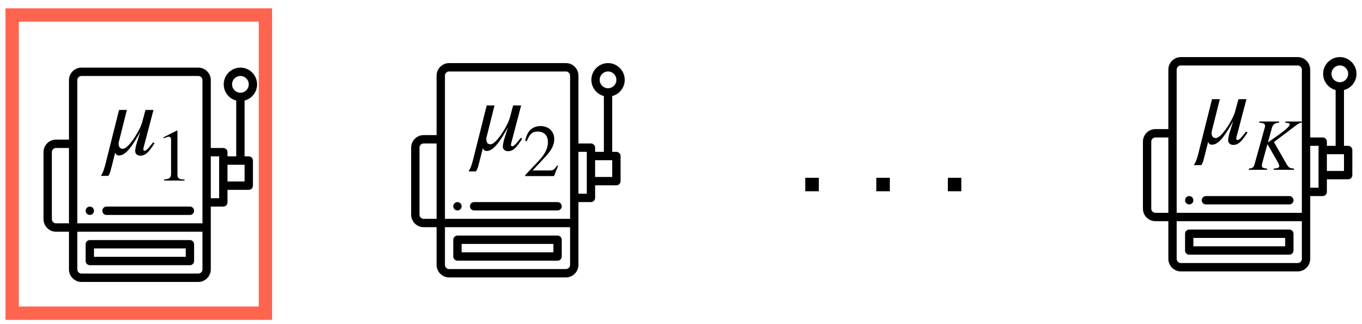


$t = 1$ $\qquad Y_1$

$t = 2$

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

Time

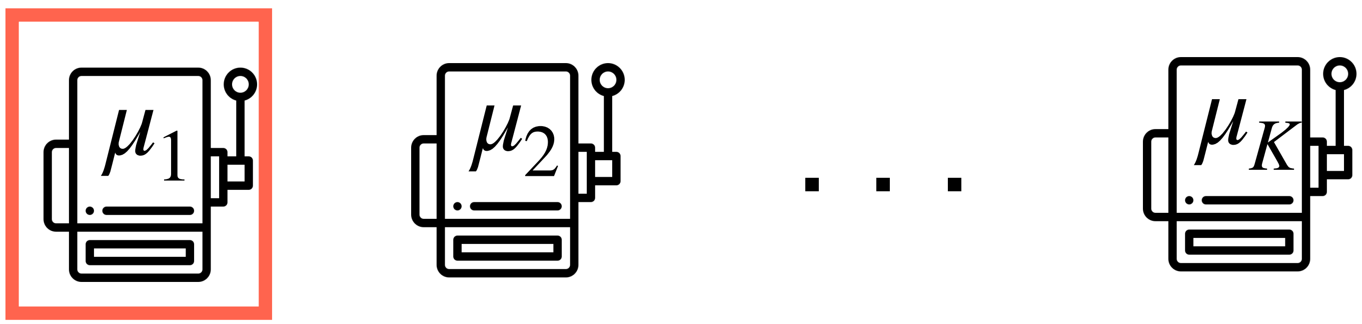


$t = 1$ $\quad Y_1$

$t = 2$ $\quad Y_2$

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

Time



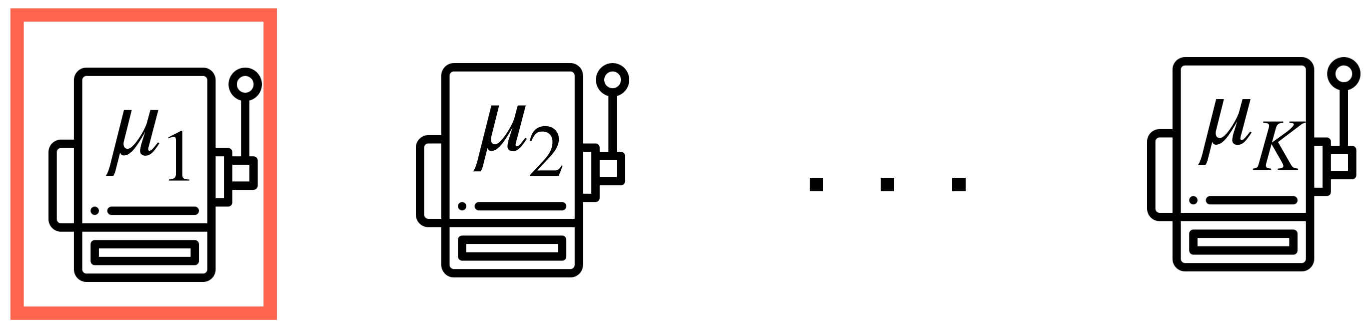
$t = 1$ Y_1

$t = 2$ Y_2

$\vdots$

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

Time



$t = 1$ $\quad Y_1$

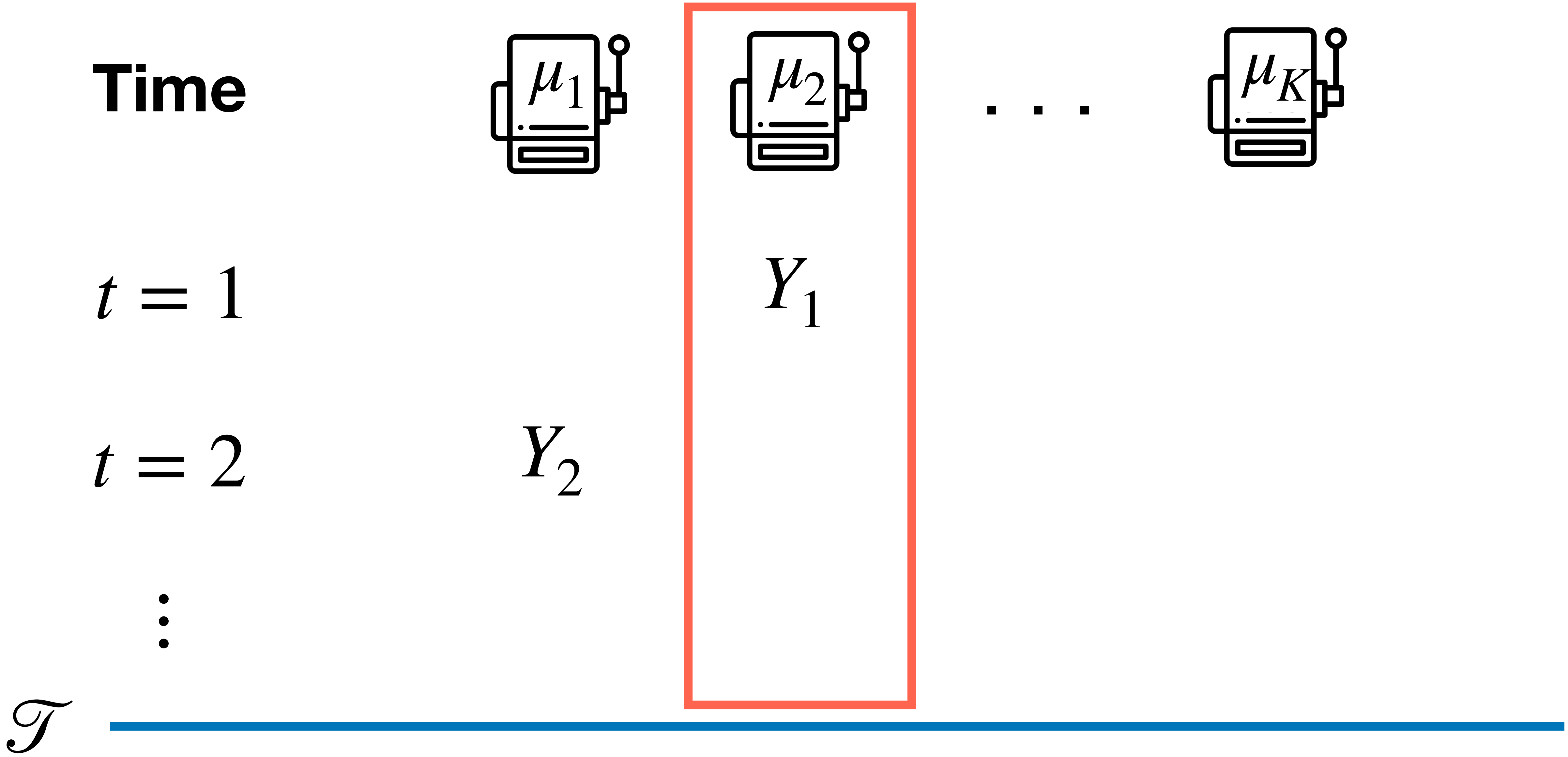
$t = 2$ $\quad Y_2$

$\vdots$

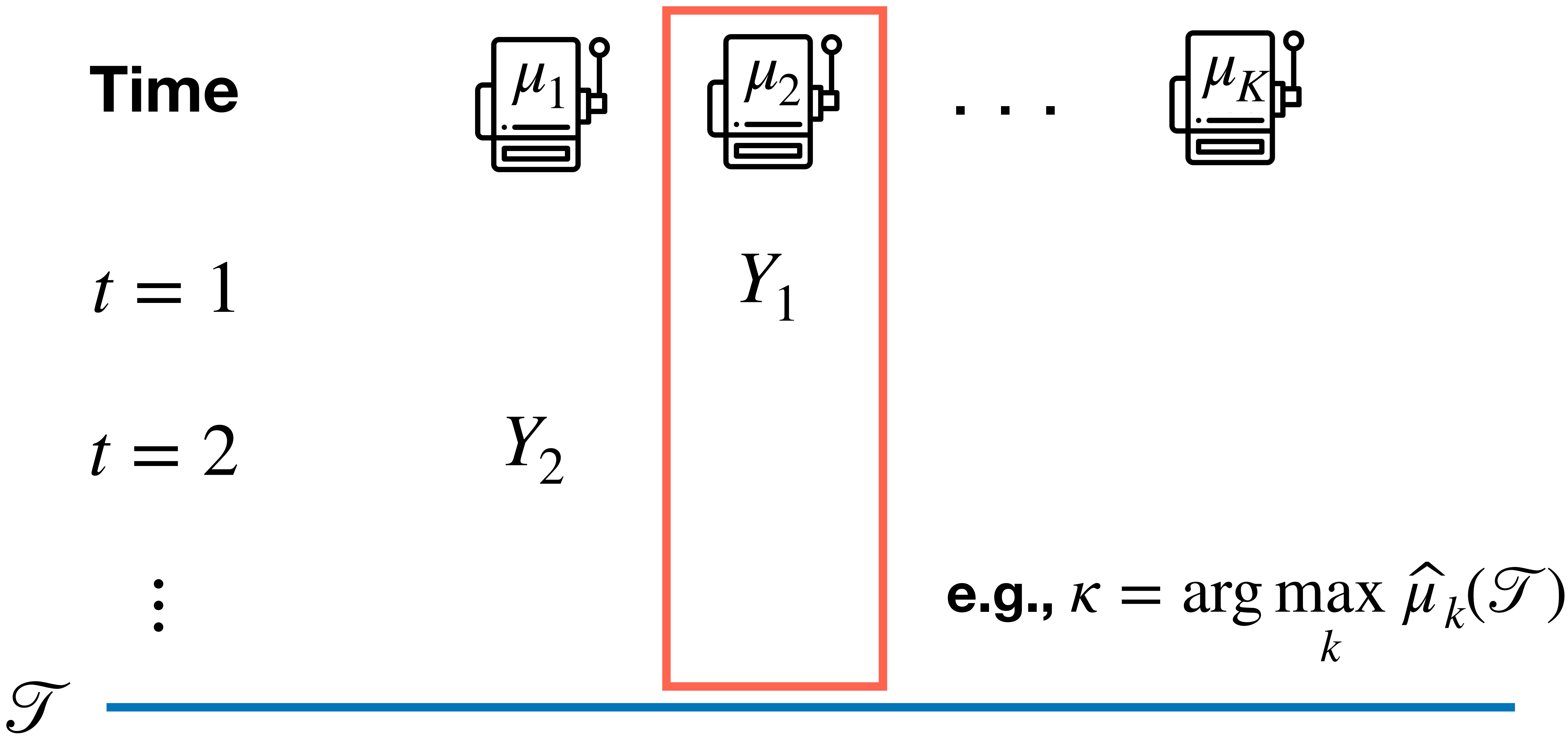
$\mathcal{T}$

Stopping time

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



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



...and the data can be used to conduct statistical inferences.

Time $\quad \mu_1 \quad \mu_2 \quad \cdots \quad \mu_K$

$t = 1 \qquad\qquad Y_1$

$t = 2 \qquad Y_2$

$\vdots$

$\mathcal{T}$

$\boxed{\widehat{\mu}_{\kappa}(\mathcal{T})}$ **Sample mean at a stopping time** $\mathcal{T}$

Q. Sign of the bias of sample mean?

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

Xu et al. [2013] :

An informal argument why the sample mean is *negatively* biased for "optimistic" algorithms.

Villar et al. [2015] :

Demonstrate this negative bias in a simulation study motivated by using MAB for clinical trials.

Q. Sign of the bias of sample mean?

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

Xu et al. [2013] :

An informal argument why the sample mean is *negatively* biased for “optimistic” algorithms.

Villar et al. [2015] :

Demonstrate this negative bias in a simulation study motivated by using MAB for clinical trials.

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**

for MABs designed to maximize cumulative reward.

Shin et al. [2019]

Introduced "monotonicity property" characterizing the bias of the sample mean for more general classes of MABs.

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

Chosen Arm **Stopping 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**

for MABs designed to maximize cumulative reward.

Shin et al. [2019]
Introduced "monotonicity property" characterizing the bias of the sample mean for more general classes of MABs.

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

Chosen Arm **Stopping Time**

However, current understanding of bias is limited in two aspects.

1. **Existing results concern the bias of the sample mean only.**

However, current understanding of bias is limited in two aspects.

1. **Existing results concern the bias of the sample mean only.**

➡ **We study the bias of monotone functions of the rewards.**

However, current understanding of bias is limited in two aspects.

1. Existing results concern the bias of the sample mean only.

➡ We study the bias of monotone functions of the rewards.

2. Existing guarantees cover only the marginal bias.

However, current understanding of bias is limited in two aspects.

1. Existing results concern the bias of the sample mean only.

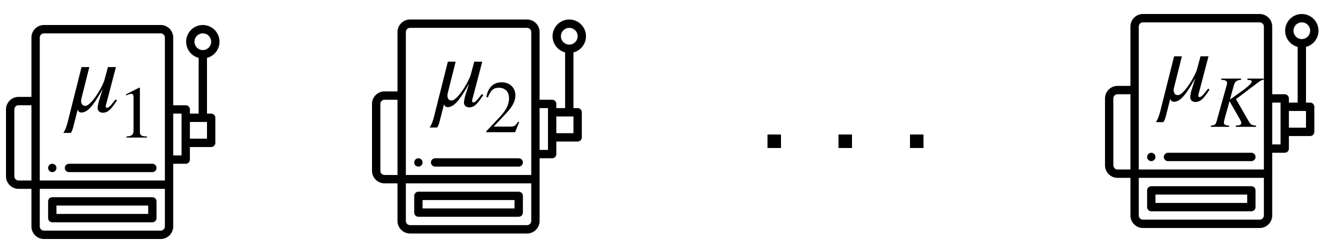
➡ We study the bias of monotone functions of the rewards.

2. Existing guarantees cover only the marginal bias.

➡ We extend previous results to cover the conditional bias.

Marginal vs Conditional bias

- **K prototype items**

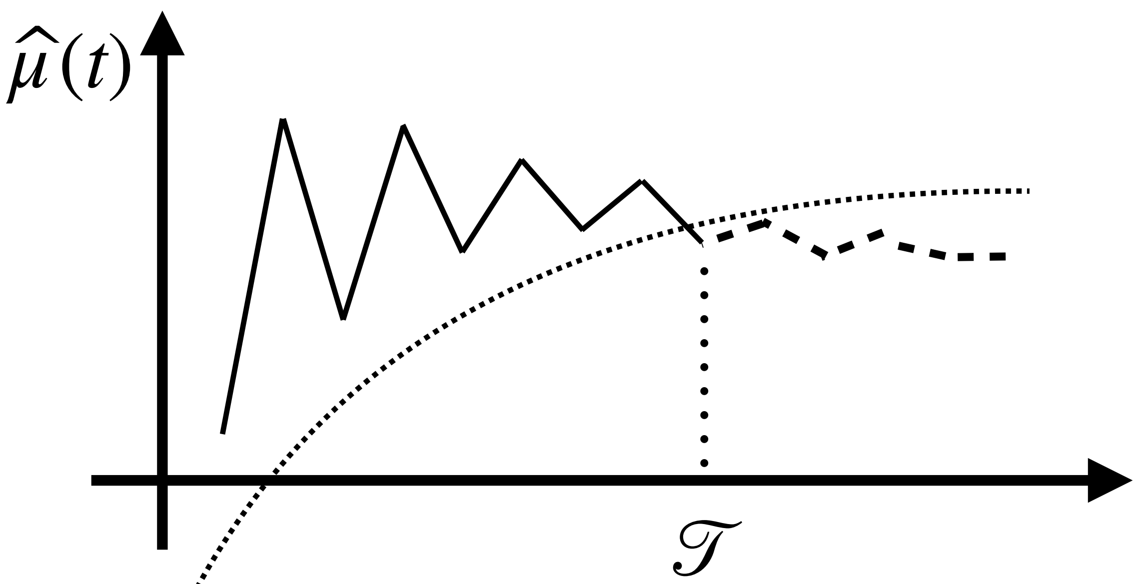


Want to screen out some items by testing

$$H_{0k} : \mu_k \geq c \quad \textbf{vs} \quad H_{1k} : \mu_k < c \ \ \textbf{for}\ k = 1, \ldots, K\,.$$

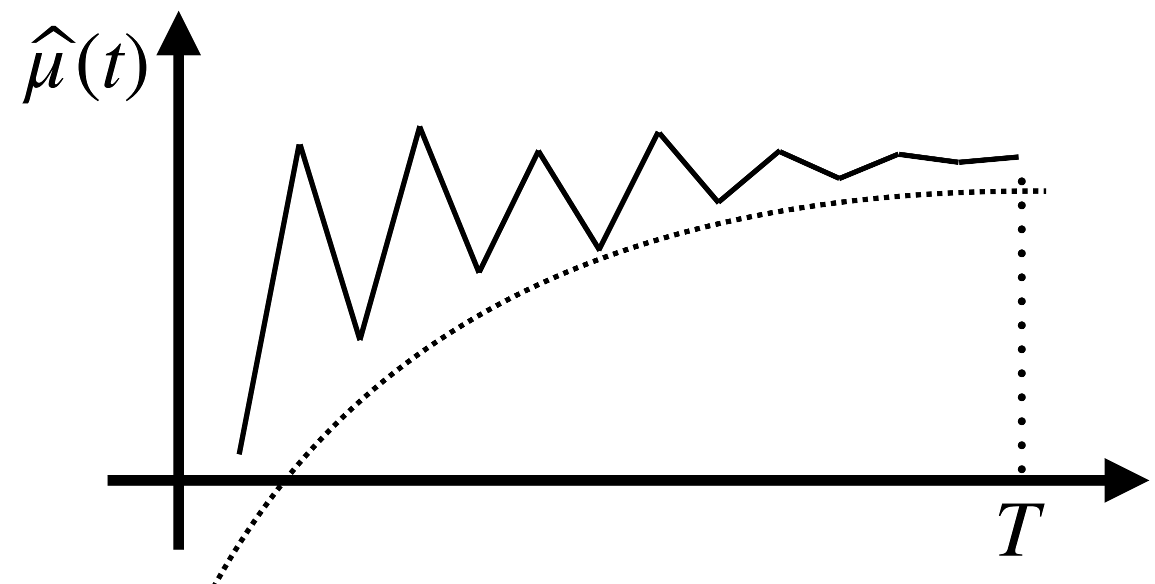
Marginal vs Conditional bias

$H_0 : \mu \geq c$ **vs** $H_1 : \mu < c$



"Screen out the item at $\mathscr{T}$ "

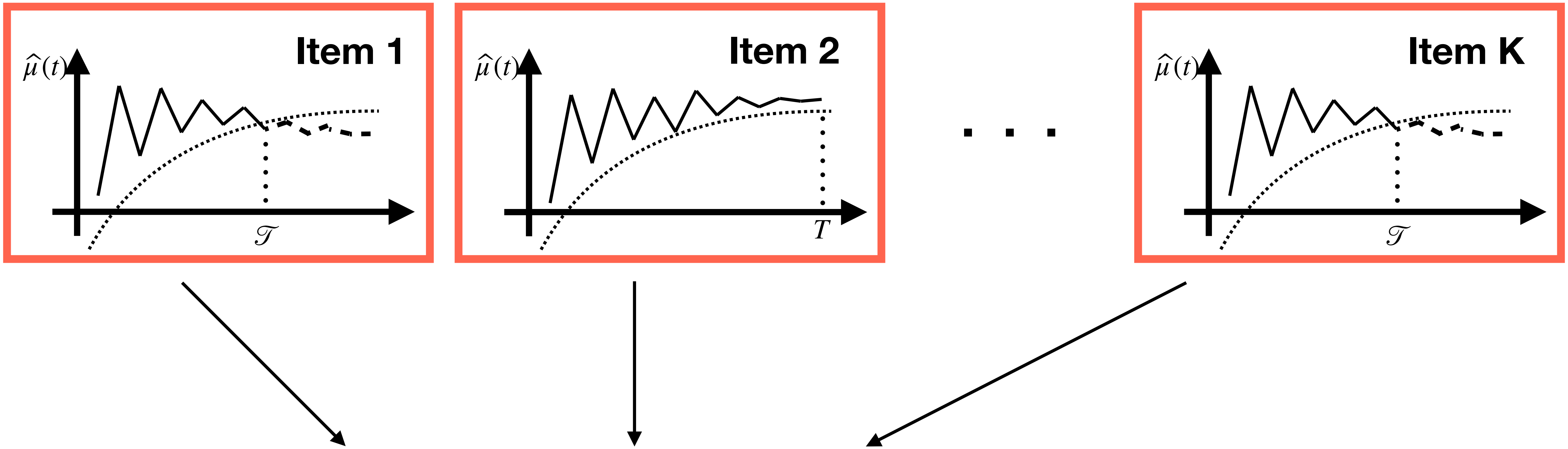
(Reject the null)



"Keep the item"

(Fail to reject the null)

Marginally,
the sample mean is negatively biased.

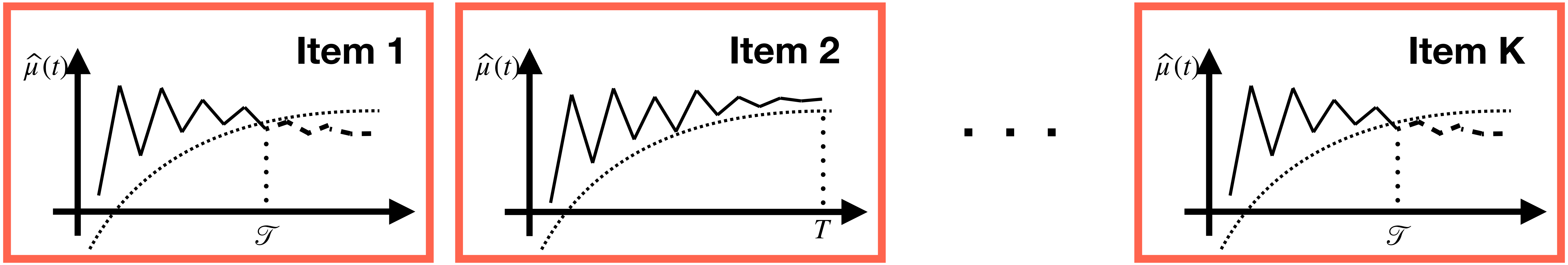


$$\mathbb{E}\left[\widehat{\mu}_k - \mu_k\right] \leq 0, \quad k = 1, \ldots, K$$

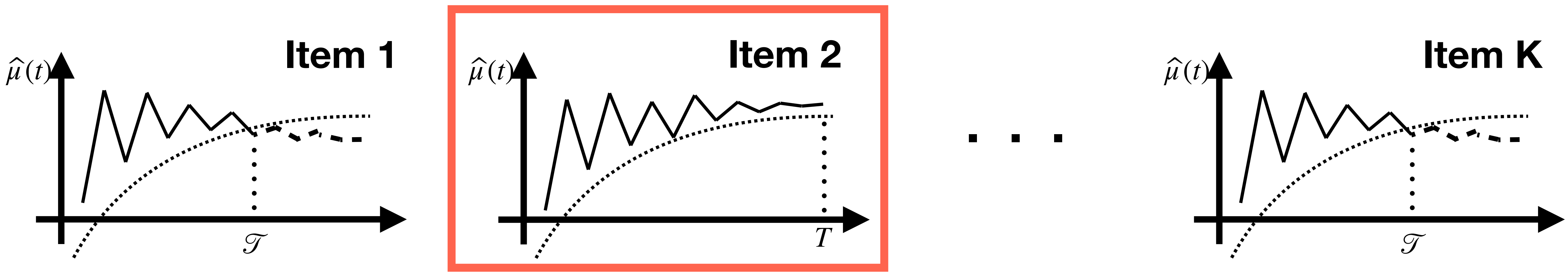
"**Underestimating** the true mean revenue."

(e.g. Starr & Woodroofe [1968], Shin et al. [2019])

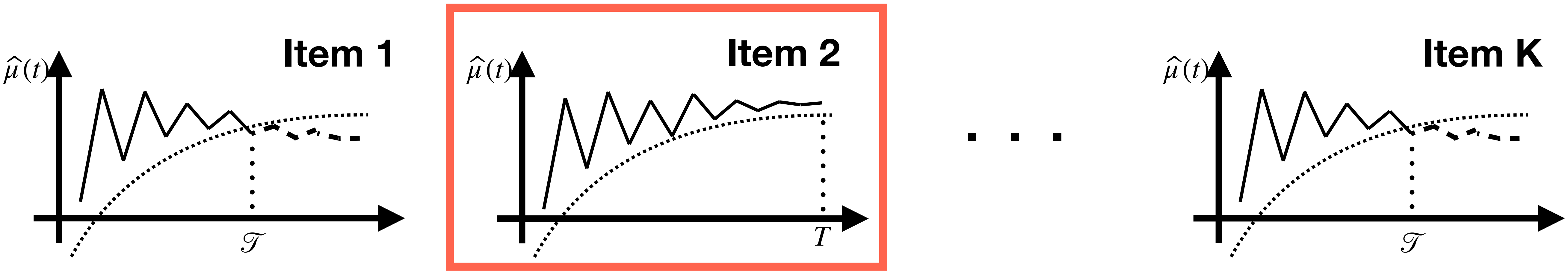
...however, we usually do not evaluate the sample mean every time.



...however, we usually do not evaluate the sample mean every time.



Conditioned on the "active" event, the sample mean is positively biased.



$$\mathbb{E}\left[\widehat{\mu}_k - \mu_k \mid \textbf{item } k \textbf{ is active}\right] \geq 0, \quad k = 1, \ldots, K$$

"**Overestimating** the true mean revenue."

Conditional bias of the empirical cumulative distribution function (CDF)

For a fixed $y \in \mathbb{R}$,

$$\mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y) \mid C\right] \leq \textbf{ or } \geq 0?$$

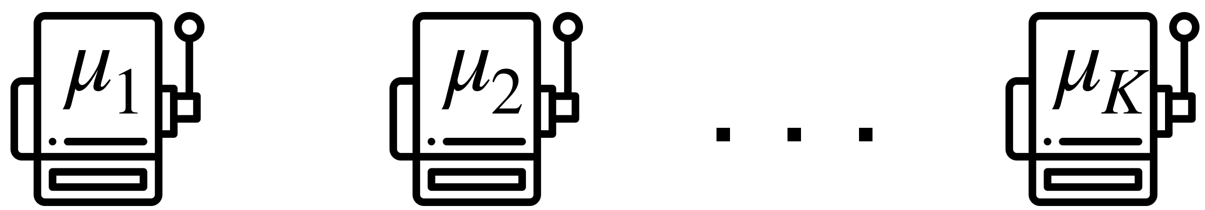
e.g., $C = \{$**Reject the null**$\}, \{$**Chosen as the best arm**$\} \ldots.$

where

$\widehat{F}_{k,\mathcal{T}}$: **Empirical CDF of arm** k **at time** $\mathcal{T}$

F_k : **True CDF of arm** k **at time** $\mathcal{T}$

Tabular model of MABs

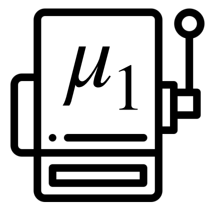


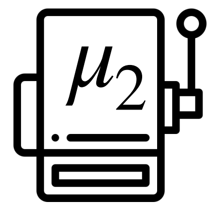
$$
\begin{array}{ll}
X^*_\infty \in \mathbb{R}^{\mathbb{N}\times K} & \begin{array}{cccc} \wr\ \textbf{i.i.d.} & \wr\ \textbf{i.i.d.} & & \wr\ \textbf{i.i.d.} \end{array} \\
:= & \left\{ \begin{array}{cccc} X^*_{1,1} & X^*_{1,2} & \cdots & X^*_{1,K} \\ X^*_{2,1} & X^*_{2,2} & \cdots & X^*_{2,K} \\ \vdots & \vdots & \vdots & \vdots \end{array} \right.
\end{array}
$$

"Hypothetical table"

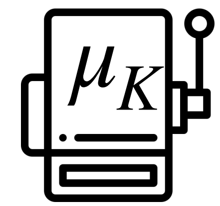
Tabular model of MABs

Time





$\cdots$



$$X^*_{1,1} \quad X^*_{1,2} \quad \cdots \quad X^*_{1,K}$$

$$X^*_{2,1} \quad X^*_{2,2} \quad \cdots \quad X^*_{2,K}$$

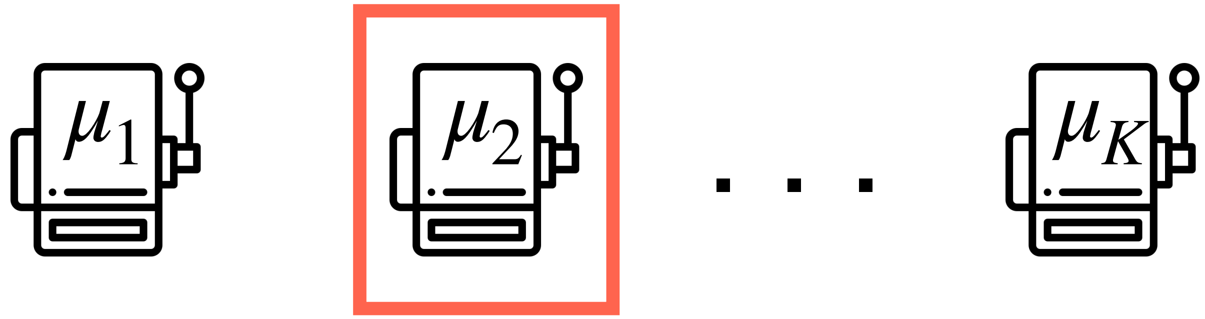
$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

Tabular model of MABs

Time	μ_1	$\boxed{\mu_2}$	$\dots$	μ_K
$t = 1$	$X^*_{1,1}$	$X^*_{1,2}$	$\dots$	$X^*_{1,K}$
	$X^*_{2,1}$	$X^*_{2,2}$	$\dots$	$X^*_{2,K}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabular model of MABs

Time

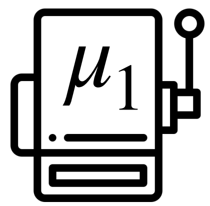
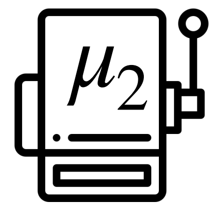
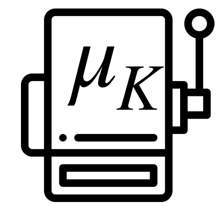


$t = 1$	$X^*_{1,1}$	Y_1	$\ldots$	$X^*_{1,K}$
	$X^*_{2,1}$	$X^*_{2,2}$	$\ldots$	$X^*_{2,K}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabular model of MABs

Time	μ_1	μ_2	$\dots$	μ_K
$t = 1$	$X^*_{1,1}$	Y_1	$\dots$	$X^*_{1,K}$
$t = 2$	$X^*_{2,1}$	$X^*_{2,2}$	$\dots$	$X^*_{2,K}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabular model of MABs

Time	 μ_1	 μ_2	$\dots$	 μ_K
$t = 1$	$X^*_{1,1}$	Y_1	$\dots$	$X^*_{1,K}$
$t = 2$	$X^*_{2,1}$	$X^*_{2,2}$	$\dots$	$X^*_{2,K}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabular model of MABs

Time	$\boxed{\mu_1}$	μ_2	$\dots$	μ_K
$t = 1$	$X^*_{1,1}$	Y_1	$\dots$	$X^*_{1,K}$
$t = 2$	Y_2	$X^*_{2,2}$	$\dots$	$X^*_{2,K}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Hypothetical dataset

Hypothetical table

$$\mathscr{D}^*_\infty = \boxed{X^*_\infty} \cup \boxed{\{W_t\}_{t=1}^{\infty}}$$

Random seeds

Hypothetical dataset

Given $\mathscr{D}^*_\infty = X^*_\infty \cup \{W_t\}_{t=1}^{\infty}$

➡ C**,** $\mathscr{T}$ **and** $N_k(t)$ **for each** t **and** k **can be expressed as some functions of** $\mathscr{D}^*_\infty$**.**

Monotone effect of a sample

Theorem

Suppose arm k has a finite mean. If $\dfrac{\mathbf{1}(C)}{N_k(\mathcal{T})}$ is an increasing function of each $X^*_{i,k}$ while keeping all other entries in $\mathcal{D}^*_\infty$ fixed then we have

$$\mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y) \mid C\right] \leq 0$$

(Negative conditional bias of the empirical CDF)

Monotone effect of a sample

<u>**Theorem**</u>

Suppose arm k has a finite mean. If $\dfrac{\mathbf{1}(C)}{N_k(\mathcal{T})}$ is an increasing function of each $X^*_{i,k}$ while keeping all other entries in $\mathcal{D}^*_\infty$ fixed then we have

$$\mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y) \mid C\right] \leq 0$$

(Negative conditional bias of the empirical CDF)

$$\mathbb{E}\left[\widehat{\mu}_k(\mathcal{T}) - \mu_k \mid C\right] \geq 0$$

(Positive conditional bias of the sample mean)

Monotone effect of a sample

Theorem

Suppose arm k has a finite mean. If $\dfrac{\mathbf{1}(C)}{N_k(\mathcal{T})}$ is a decreasing function of each $X^*_{i,k}$ while keeping all other entries in $\mathcal{D}^*_\infty$ fixed then we have

$$\mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y) \mid C\right] \geq 0$$

(Positive conditional bias of the empirical CDF)

Monotone effect of a sample

Theorem

Suppose arm k has a finite mean. If $\dfrac{\mathbf{1}(C)}{N_k(\mathcal{T})}$ is a decreasing function of each $X^*_{i,k}$ while keeping all other entries in $\mathcal{D}^*_\infty$ fixed then we have

$$\mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y) \mid C\right] \geq 0$$

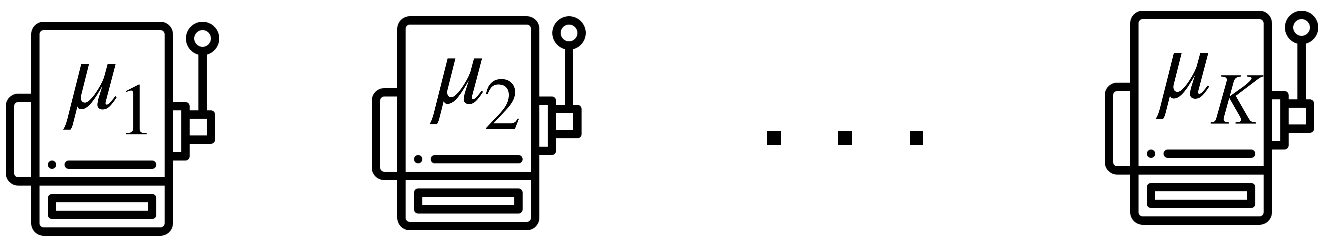
(Positive conditional bias of the empirical CDF)

$$\mathbb{E}\left[\widehat{\mu}_k(\mathcal{T}) - \mu_k \mid C\right] \leq 0$$

(Negative conditional bias of the sample mean)

E.g.: Best arm identification

- **K prototype items**

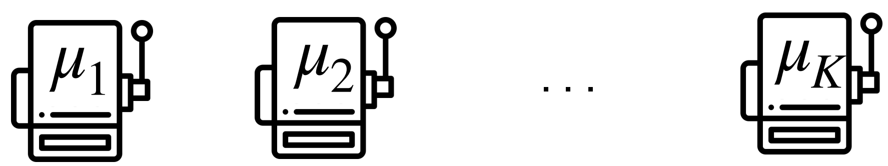


Want to figure out which one has the largest revenue.

E.g.: Best arm identification

lil' UCB algorithm

Time



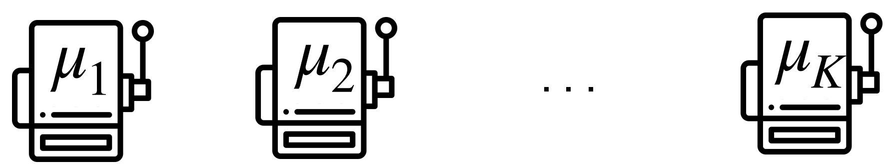
$t = 1$ $\qquad$ Y_1

$t = 2$ $\qquad$ Y_2

E.g.: Best arm identification

lil' UCB algorithm

Time



$t = 1$ $\quad\quad\quad\quad\quad\quad\quad\quad$ Y_1

$t = 2$ $\quad\quad\quad\quad$ Y_2

$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$ **(Upper confidence bound)**

E.g.: Best arm identification

lil' UCB algorithm

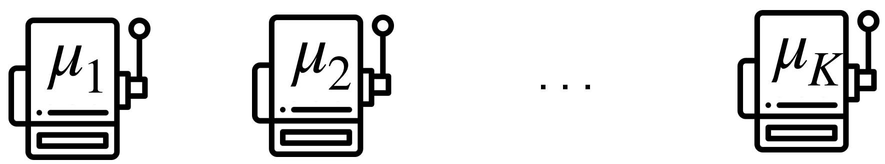
Time	μ_1	μ_2	$\cdots$	μ_K
$t = 1$		Y_1		
$t = 2$	Y_2			

$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$ **(Upper confidence bound)**

$$\mathcal{T} = \inf \left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

E.g.: Best arm identification

lil' UCB algorithm



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

$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$ **(Upper confidence bound)**

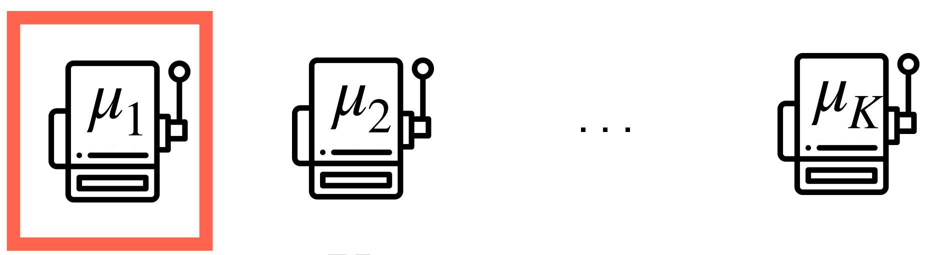
$$\mathscr{T} = \inf\left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathscr{T})$$

E.g.: Best arm identification

lil' UCB algorithm

Time



$t = 1$ $\quad$ Y_1

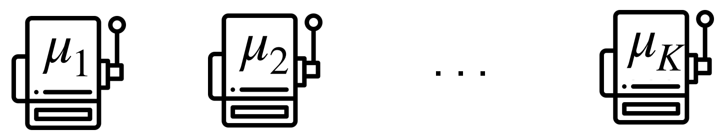
$t = 2$ $\quad$ Y_2

a) Item 1 is chosen as the best.

b) Item 1 is NOT chosen as the best.

E.g.: Best arm identification

a) Item 1 is chosen as the best ($\kappa = 1$).



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

Sample from item 1 $\xmapsto{\text{Increasing}}$ $\dfrac{\mathbf{1}(\kappa = 1)}{N_1(\mathcal{T})}$

$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$

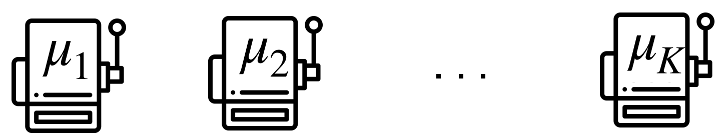
$$\mathcal{T} = \inf \left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathcal{T})$$

E.g.: Best arm identification

a) Item 1 is chosen as the best ($\kappa = 1$).

Time



$t = 1$ $\qquad\qquad Y_1$ (arm 2)

$t = 2$ $\qquad Y_2$ (arm 1)

$$A_t = \arg\max_k \hat{\mu}_k(t) + u(N_k(t))$$

$$\mathcal{T} = \inf\left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathcal{T})$$

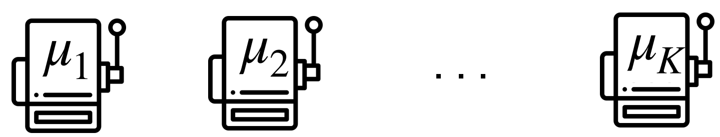
Sample from item 1 $\xrightarrow{\text{Increasing}}$ $\dfrac{\mathbf{1}(\kappa = 1)}{N_1(\mathcal{T})}$

➡ **Negative conditional bias of the empirical CDF**

E.g.: Best arm identification

a) Item 1 is chosen as the best ($\kappa = 1$).

Time



$t = 1$ $\quad Y_1$

$t = 2$ $\quad Y_2$

Sample from item 1 $\xrightarrow{\text{Increasing}}$ $\dfrac{\mathbf{1}(\kappa = 1)}{N_1(\mathscr{T})}$

$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$

$$\mathscr{T} = \inf\left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathscr{T})$$

- **Negative conditional bias of the empirical CDF**
- **Positive conditional bias of the sample mean**

E.g.: Best arm identification

b) Item 1 is <u>NOT</u> chosen as the best ($\kappa \neq 1$).

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

Sample from item 1 $\xmapsto{\text{Decreasing}}$ $\dfrac{\mathbf{1}(\kappa \neq 1)}{N_1(\mathcal{T})}$

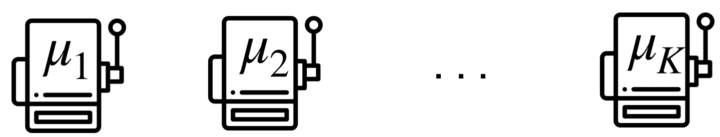
$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$

$$\mathcal{T} = \inf \left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathcal{T})$$

E.g.: Best arm identification

b) Item 1 is <u>NOT</u> chosen as the best ($\kappa \neq 1$).



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

$$A_t = \arg\max_k \widehat{\mu}_k(t) + u(N_k(t))$$

$$\mathcal{T} = \inf\left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathcal{T})$$

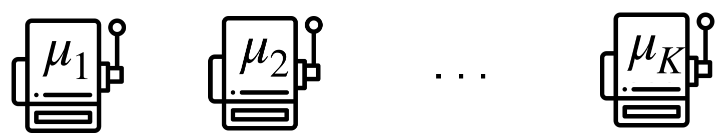
Sample from item 1 $\xmapsto{\text{Decreasing}}$ $\dfrac{\mathbf{1}(\kappa \neq 1)}{N_1(\mathcal{T})}$

➡ **Positive conditional bias of the empirical CDF**

E.g.: Best arm identification

b) Item 1 is <u>NOT</u> chosen as the best ($\kappa \neq 1$).

Time



$t = 1$ $\quad Y_1$

$t = 2$ $\quad Y_2$

Sample from item 1 $\xmapsto{\text{Decreasing}}$ $\dfrac{\mathbf{1}(\kappa \neq 1)}{N_1(\mathcal{T})}$

$$A_t = \arg\max_k \hat{\mu}_k(t) + u(N_k(t))$$

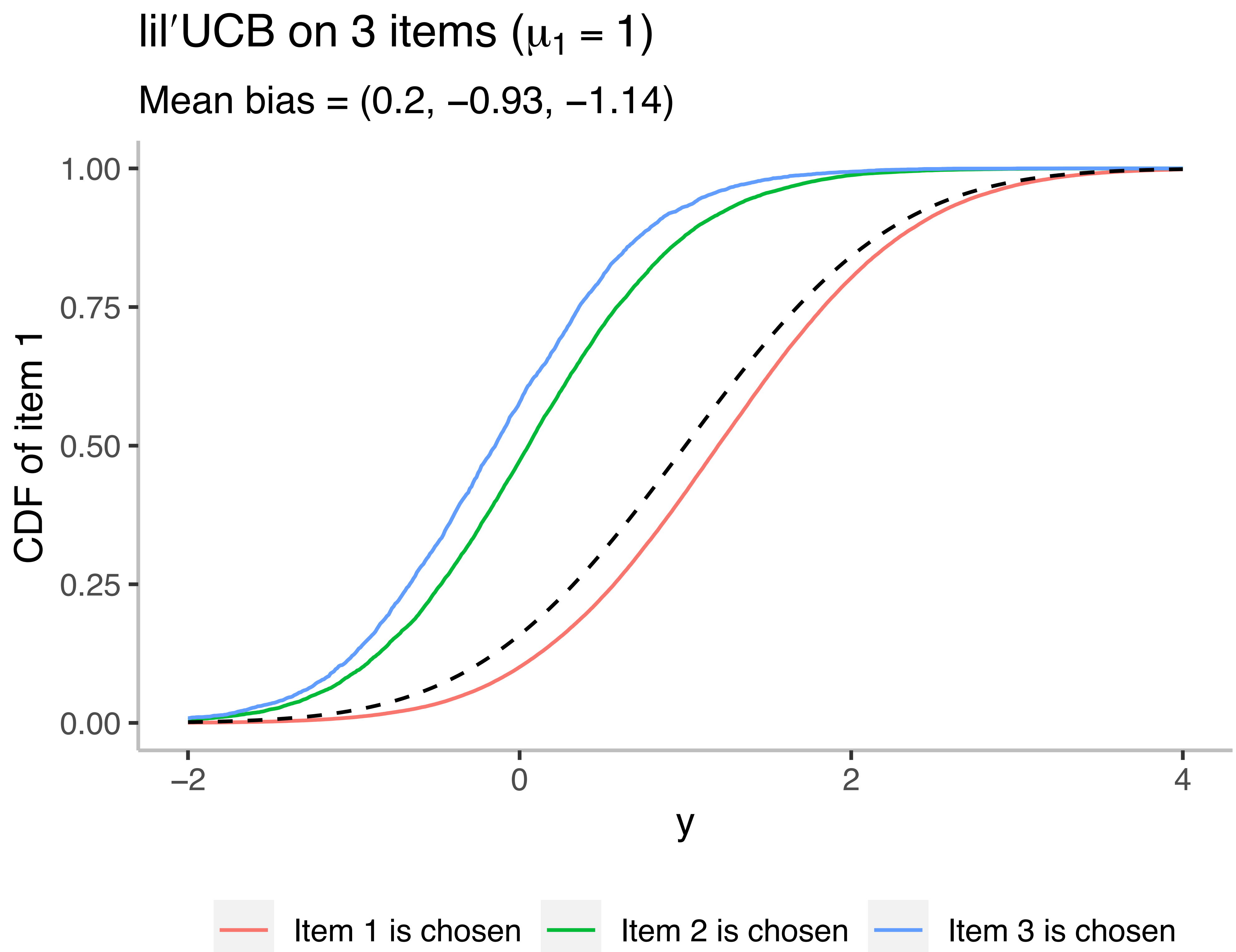
$$\mathcal{T} = \inf \left\{ t : \exists k, N_k(t) \geq 1 + \lambda \sum_{j \neq k} N_j(t) \right\}$$

$$\kappa = \arg\max_k N_k(\mathcal{T})$$

➡ **Positive conditional bias of the empirical CDF**

➡ **Negative conditional bias of the sample mean**

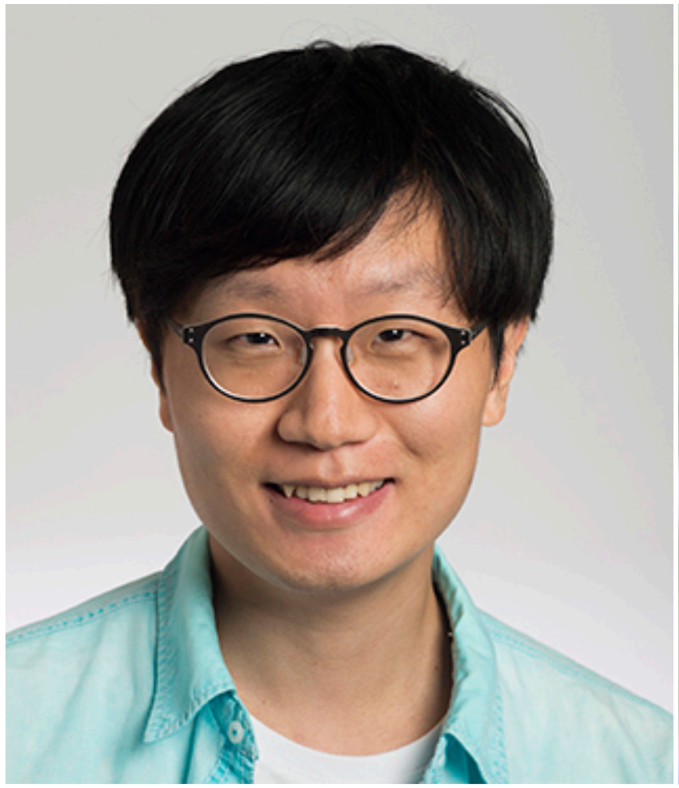
Average of the empirical CDF of item 1 conditioned on each event



Thank you!

On conditional versus marginal bias in multi-armed bandits

Jaehyeok Shin, Aaditya Ramdas and Alessandro Rinaldo



Carnegie Mellon University