BEST RESTLESS MARKOV ARM IDENTIFICATION

2022 IEEE Information Theory Workshop, Mumbai, India

P. N. Karthik
karthik@nus.edu.sg

Srinivas Reddy Kota
ksreddy@nus.edu.sg

Vincent Y. F. Tan
vtan@nus.edu.sg

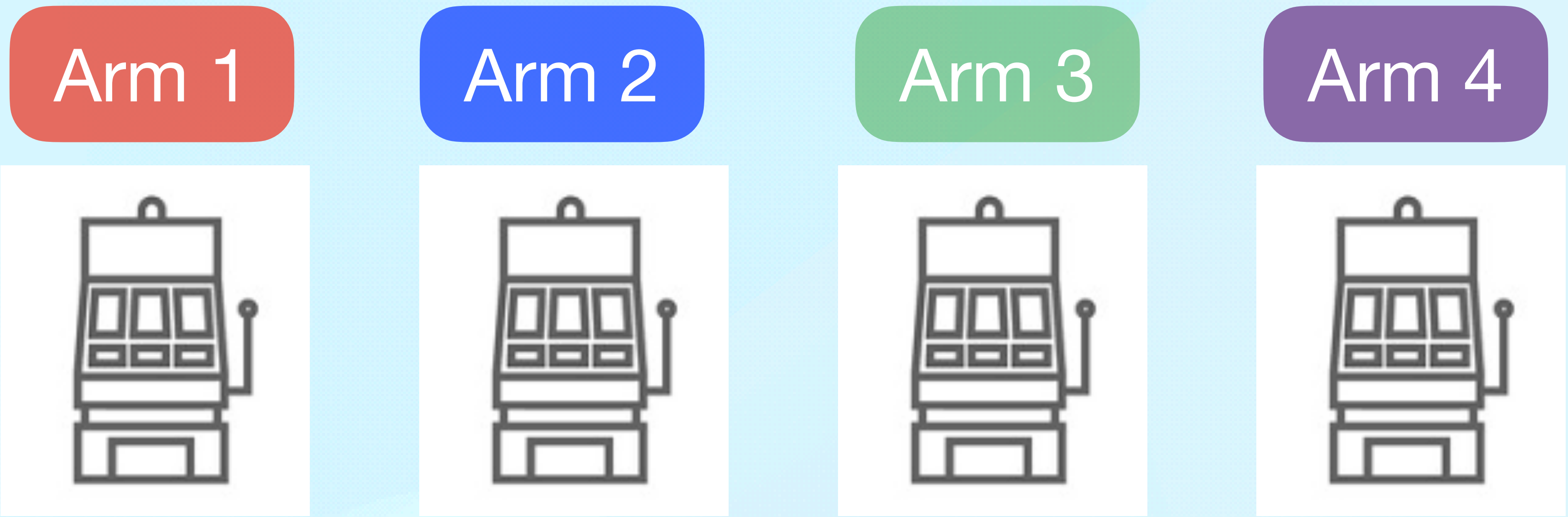
National University of Singapore

08 November 2022

NUS
National University of Singapore

PROBLEM SETUP & OBJECTIVE

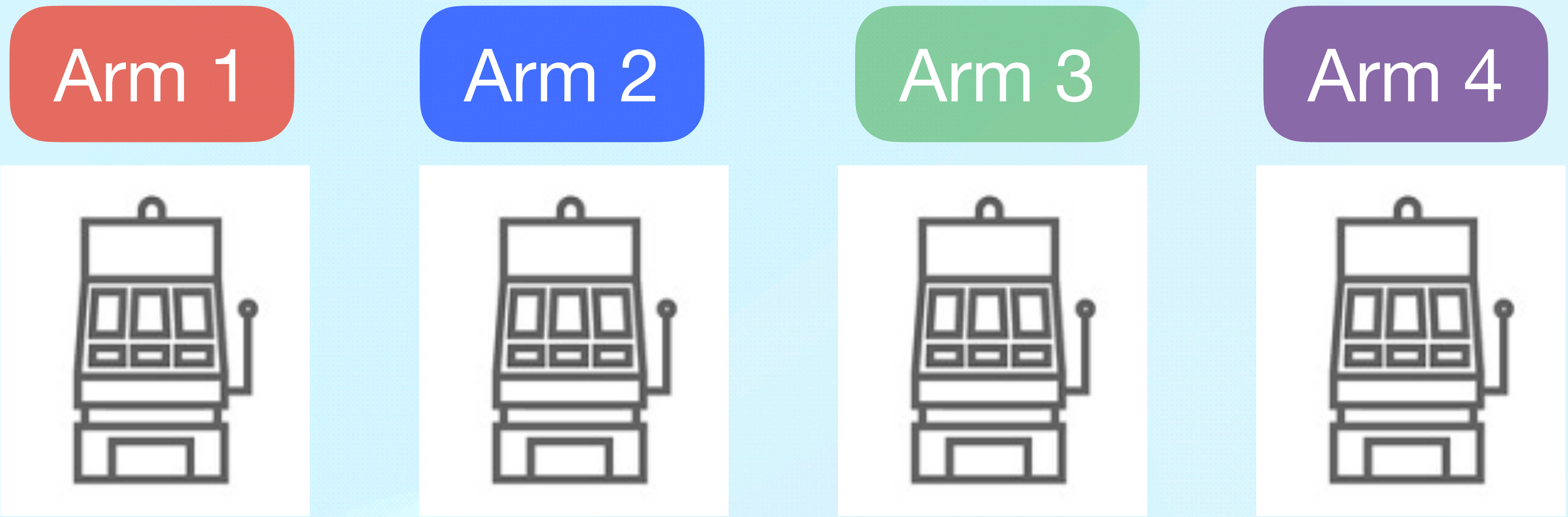
PROBLEM SETUP & OBJECTIVE



* TPM(s): transition probability matrix(ces)

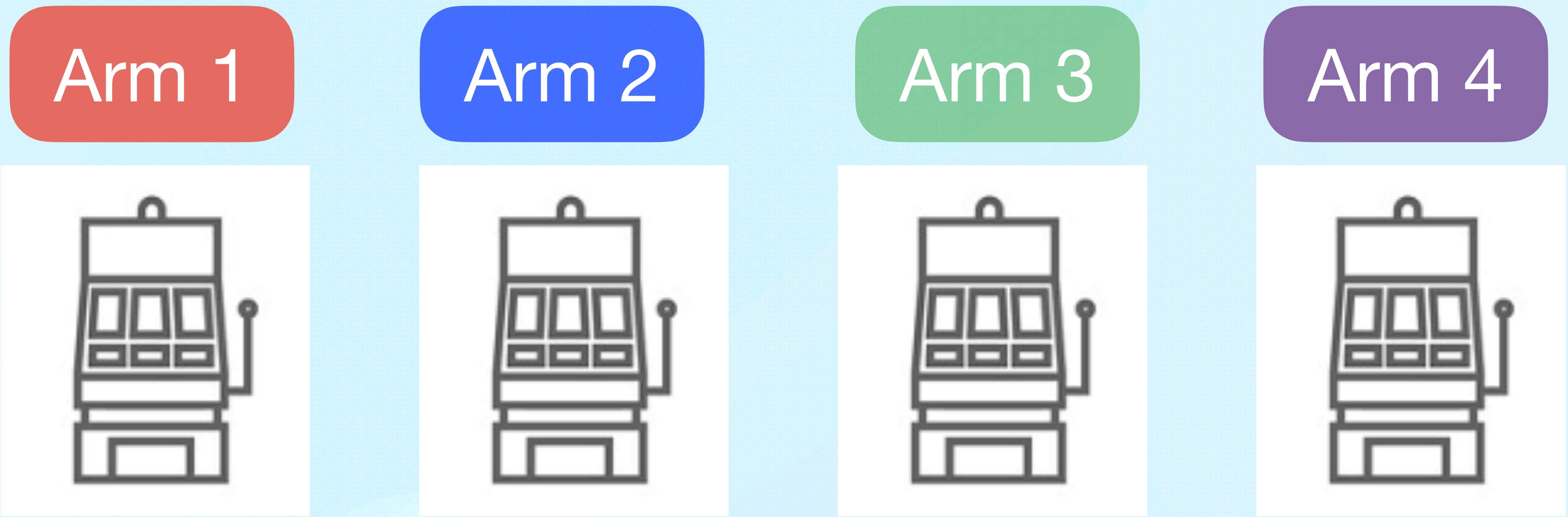
PROBLEM SETUP & OBJECTIVE

- A multi-armed bandit with $K \geq 2$ arms



* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE

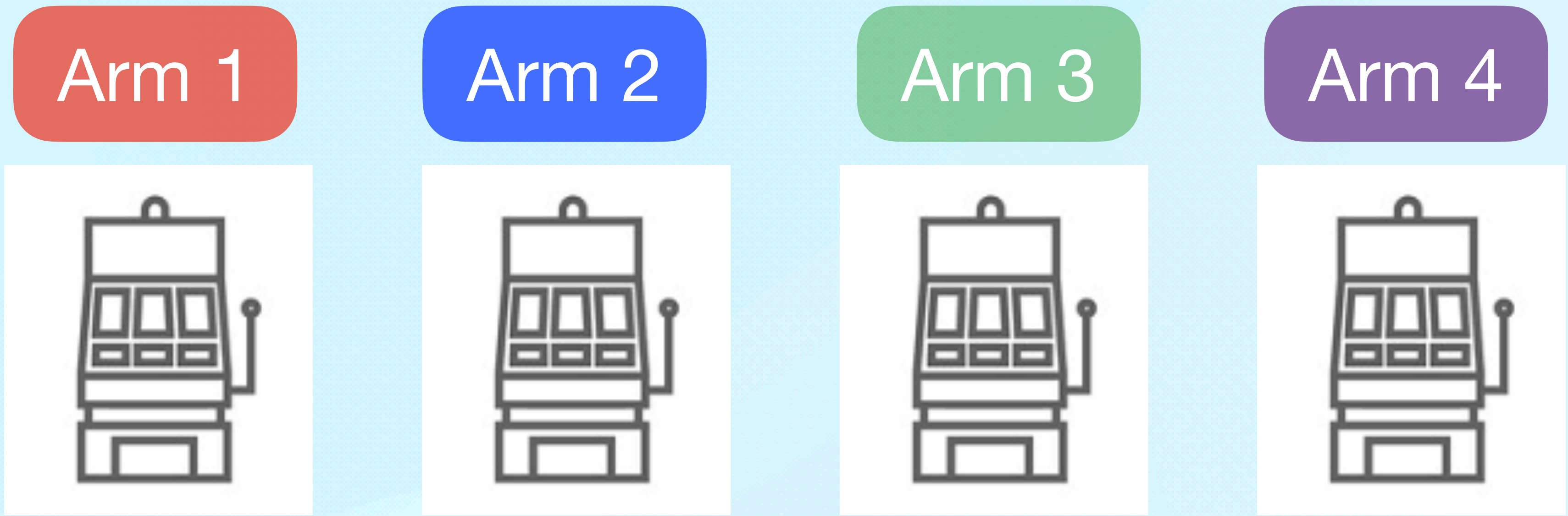


- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$

* TPM(s): transition probability matrix(ces)

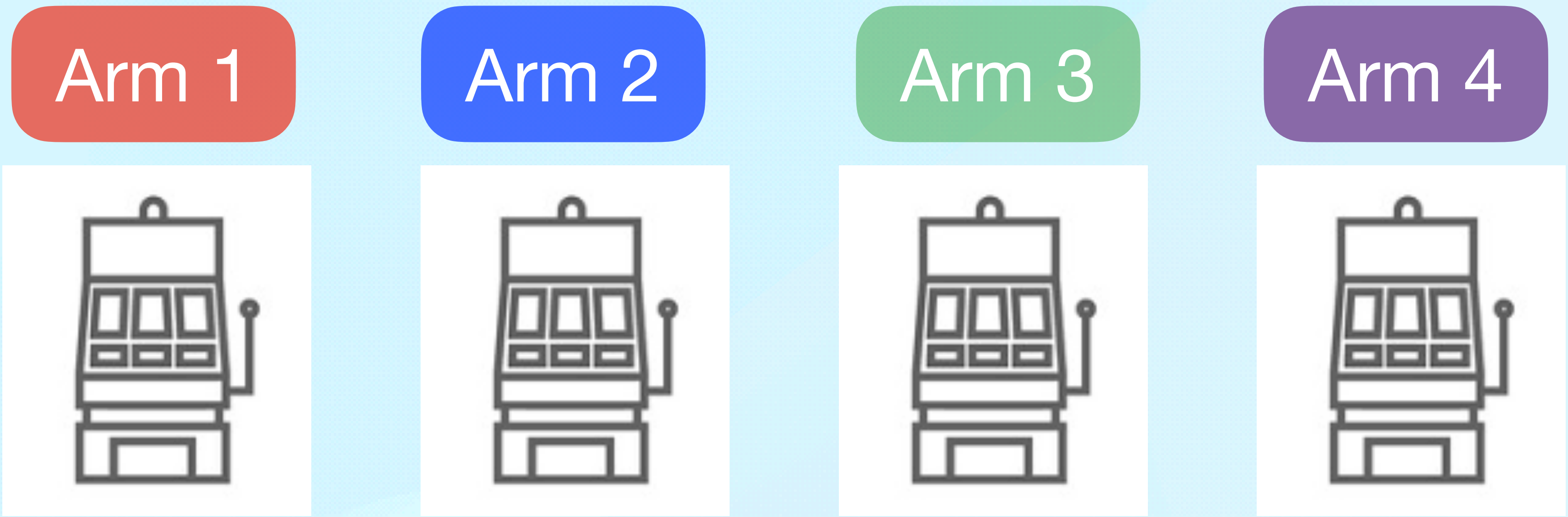
PROBLEM SETUP & OBJECTIVE

- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless



* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE



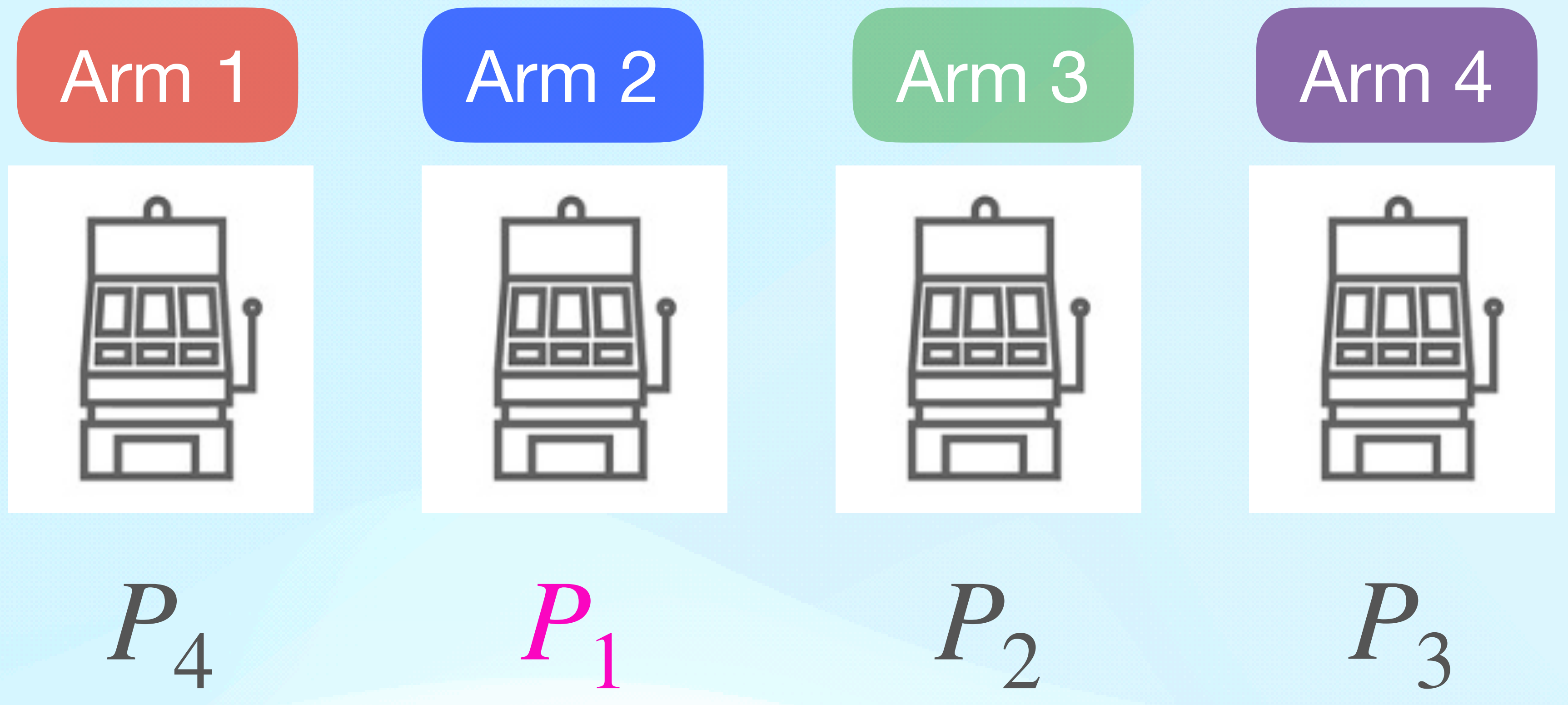
- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \rightarrow [K]$, the TPM of arm a is $P_{\sigma(a)}$

* TPM(s): transition probability matrix(ces)

NUS National University of Singapore

PROBLEM SETUP & OBJECTIVE

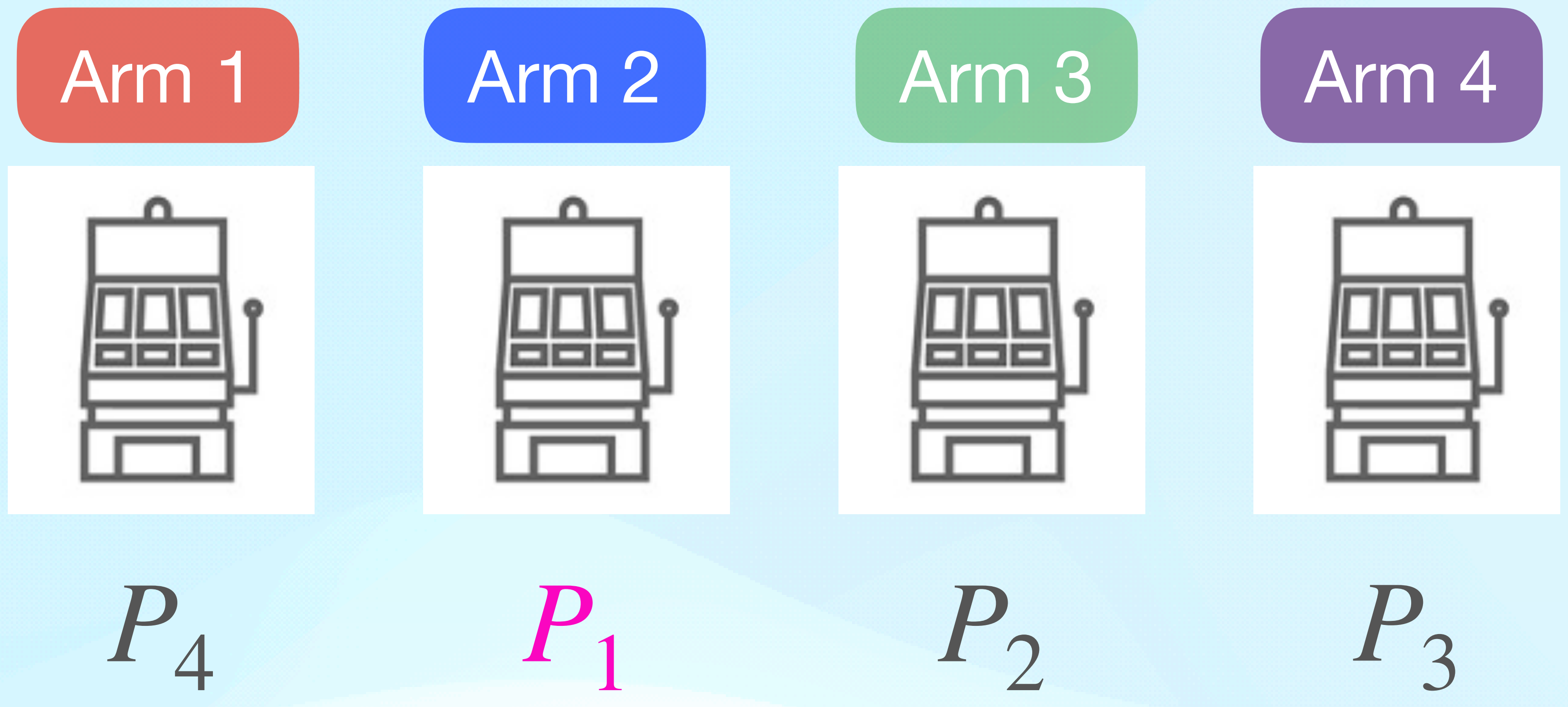
- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$



* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE

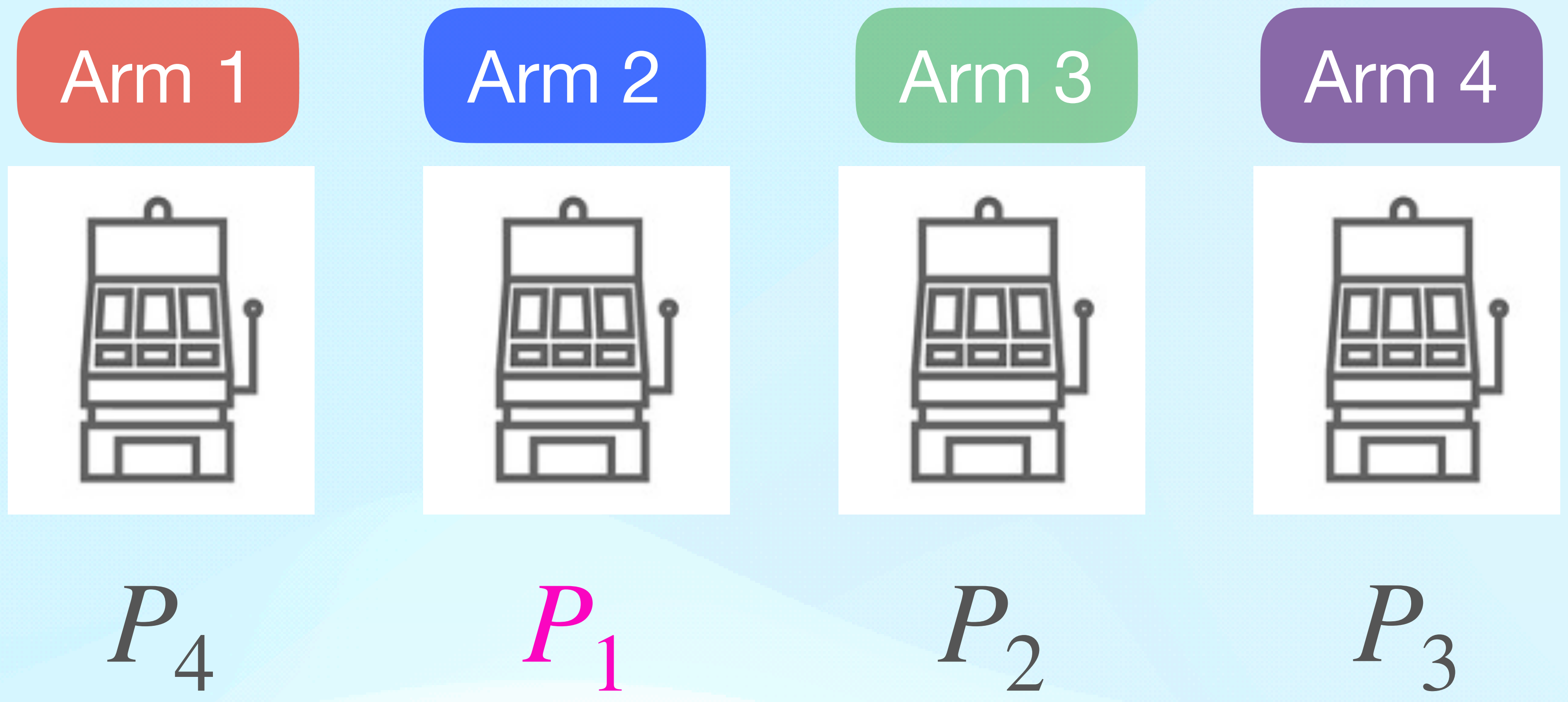
- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown



* TPM(s): transition probability matrix(ces)

NUS National University of Singapore

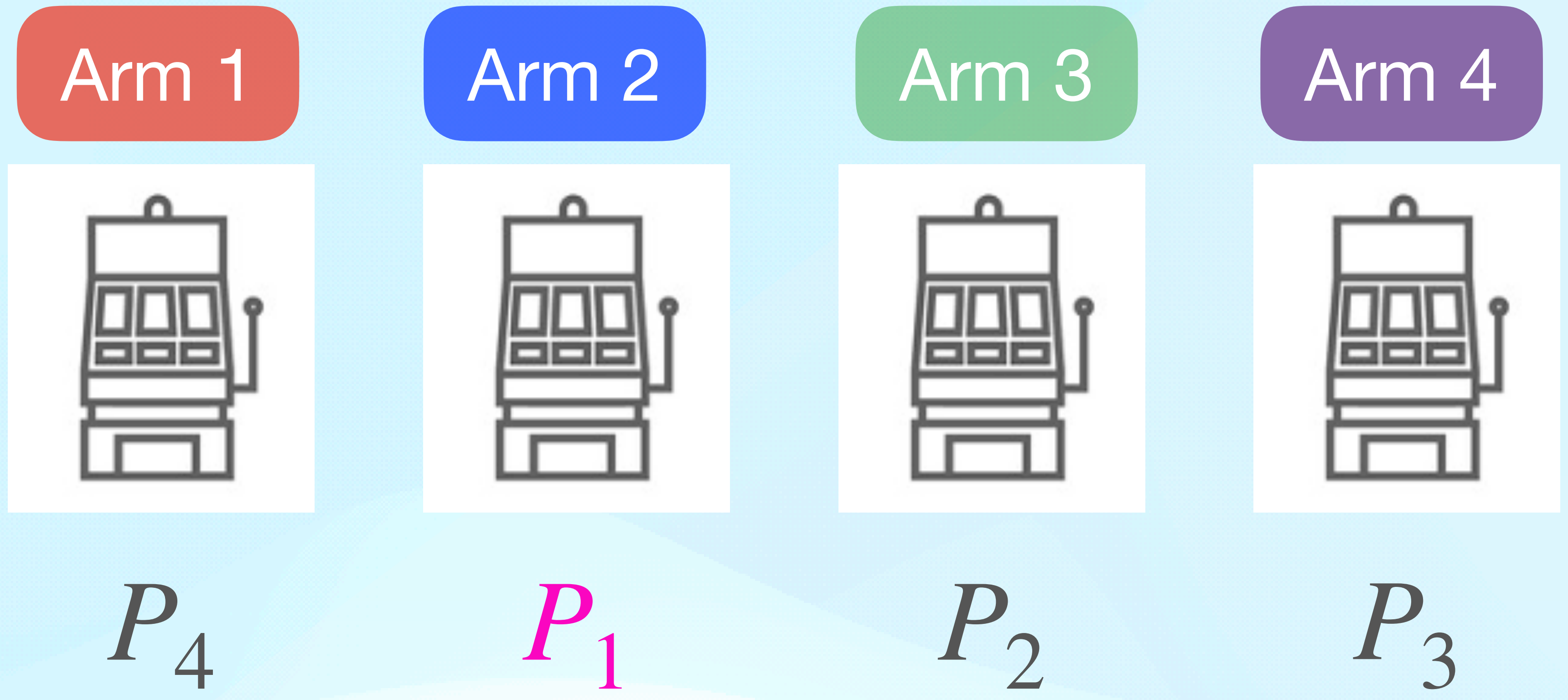
PROBLEM SETUP & OBJECTIVE



- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown
- Stationary distribution of TPM P_k is μ_k

* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE



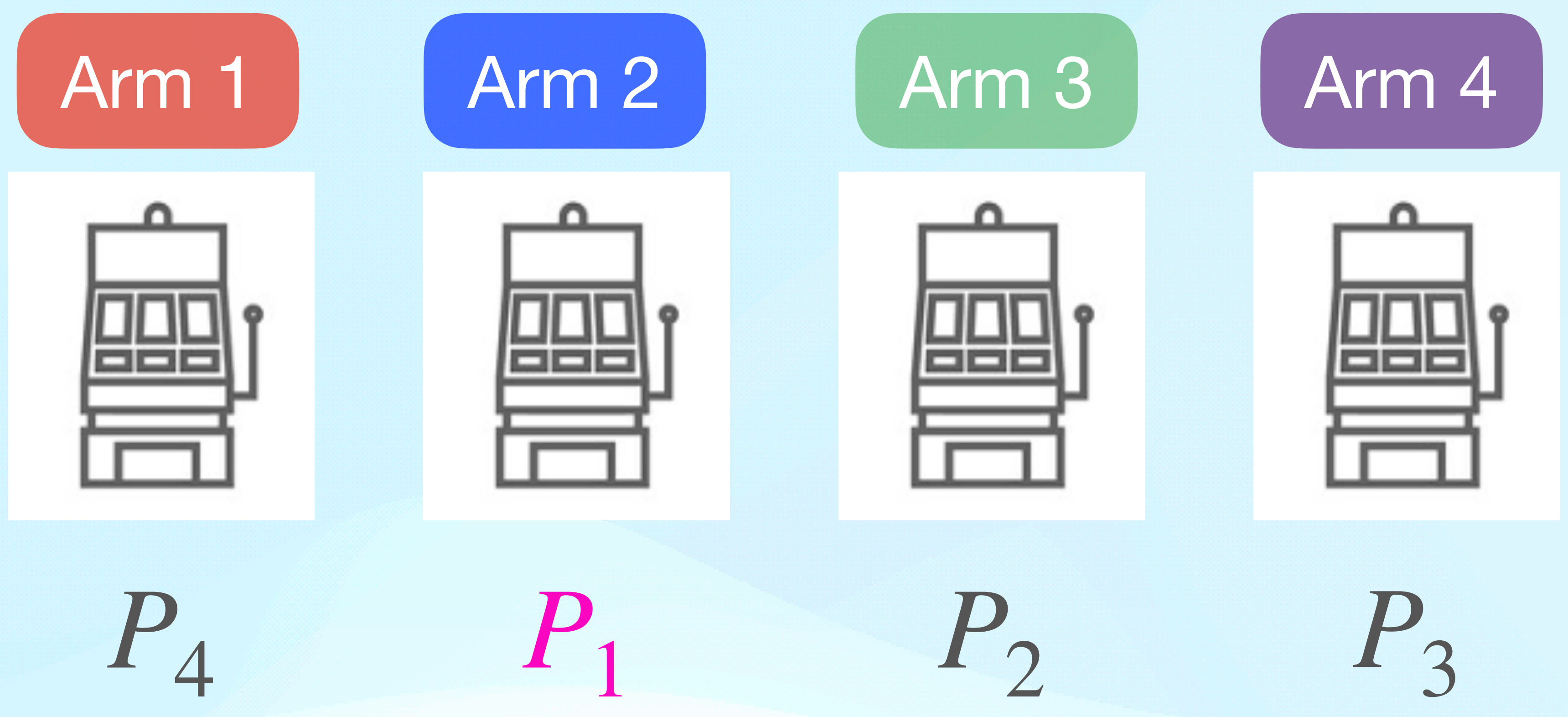
- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown
- Stationary distribution of TPM P_k is μ_k
- Given $f : \mathcal{S} \to \mathbb{R}$, the best arm

$$a^\star := \arg\max_a \langle f, \mu_{\sigma(a)} \rangle = \arg\max_a \sum_{i \in \mathcal{S}} f(i)\ \mu_{\sigma(a)}(i)$$

* TPM(s): transition probability matrix(ces)

NUS National University of Singapore

PROBLEM SETUP & OBJECTIVE



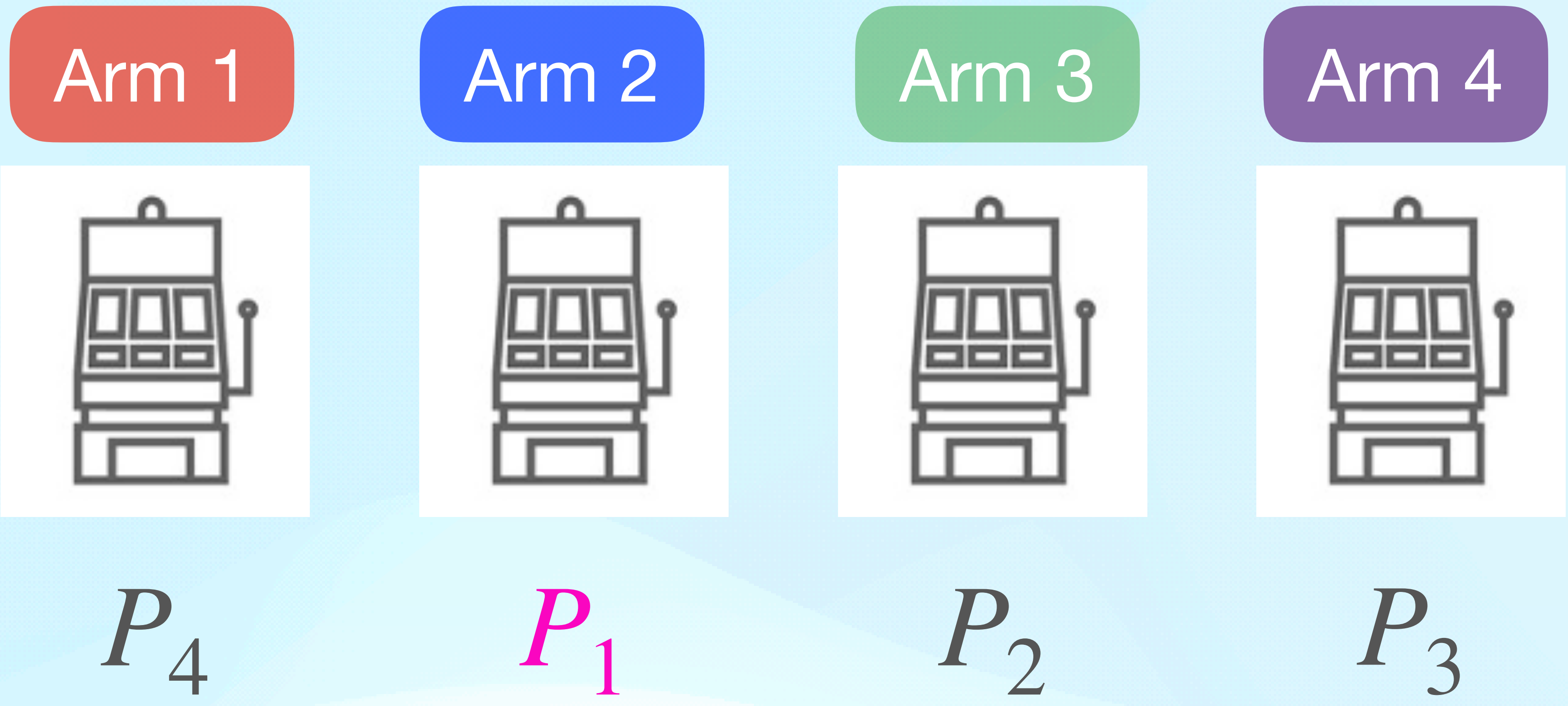
- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown
- Stationary distribution of TPM P_k is μ_k
- Given $f : \mathcal{S} \to \mathbb{R}$, the best arm

$$a^\star := \arg\max_a \langle f, \mu_{\sigma(a)} \rangle = \arg\max_a \sum_{i \in \mathcal{S}} f(i)\ \mu_{\sigma(a)}(i)$$

goal: To determine the best arm quickly and with high confidence

* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE



- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time **Markov** chain on **finite** state space $\mathcal{S}$
- Arms are **restless**
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- **TPMs** are **known**, σ is **unknown**
- Stationary distribution of TPM P_k is μ_k
- Given $f : \mathcal{S} \to \mathbb{R}$, the **best arm**

$$a^\star := \arg\max_a \langle f, \mu_{\sigma(a)} \rangle = \arg\max_a \sum_{i \in \mathcal{S}} f(i)\ \mu_{\sigma(a)}(i)$$

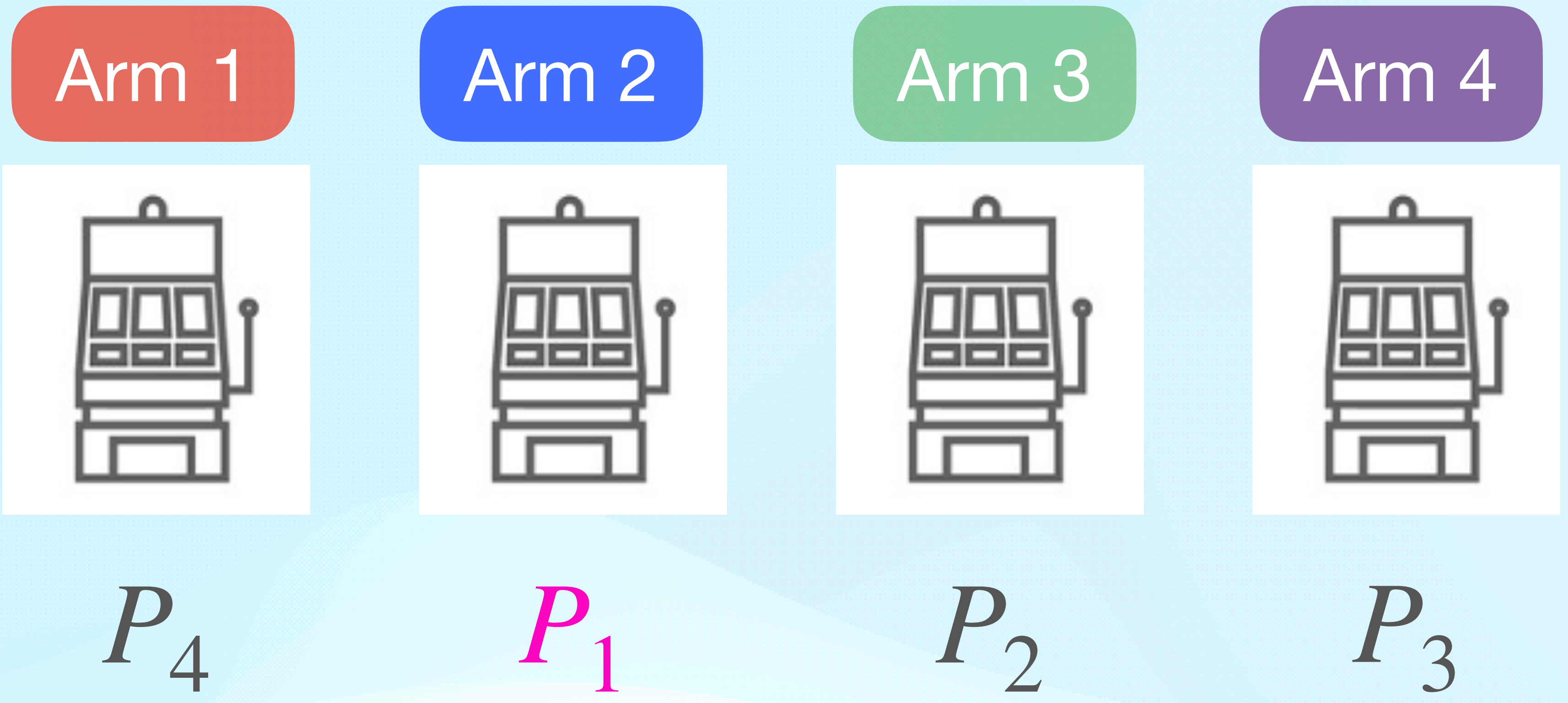
accurately $\Pi(\delta) = \{\pi : P^\pi(\hat{a} \neq a^\star) \leq \delta\}$

goal To determine the best arm quickly and with high confidence

NUS National University of Singapore

* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE



- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown
- Stationary distribution of TPM P_k is μ_k
- Given $f : \mathcal{S} \to \mathbb{R}$, the best arm

$$a^\star := \arg\max_a \langle f, \mu_{\sigma(a)} \rangle = \arg\max_a \sum_{i \in \mathcal{S}} f(i)\ \mu_{\sigma(a)}(i)$$

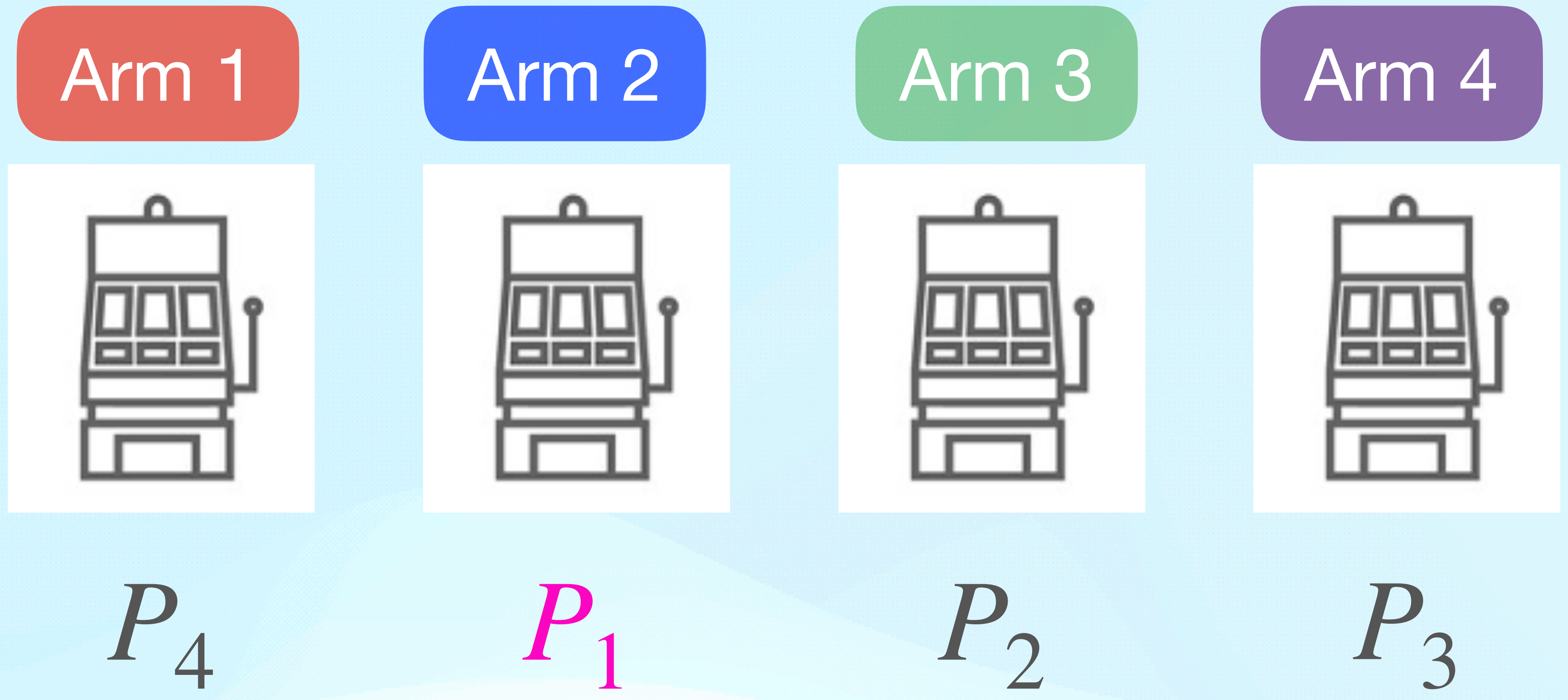
accurately $\Pi(\delta) = \{\pi : P^\pi(\hat{a} \neq a^\star) \leq \delta\}$

quickly $\inf_{\pi \in \Pi(\delta)} \mathbb{E}^\pi[\text{stopping time under } \pi]$

goal To determine the best arm quickly and with high confidence

* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE



- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \to [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown
- Stationary distribution of TPM P_k is μ_k
- Given $f : \mathcal{S} \to \mathbb{R}$, the best arm

$$a^\star := \arg\max_a \langle f, \mu_{\sigma(a)} \rangle = \arg\max_a \sum_{i \in \mathcal{S}} f(i)\ \mu_{\sigma(a)}(i)$$

accurately $\Pi(\delta) = \{\pi : P^\pi(\hat{a} \neq a^\star) \leq \delta\}$

quickly $\inf_{\pi \in \Pi(\delta)} \mathbb{E}^\pi[\text{stopping time under } \pi]$

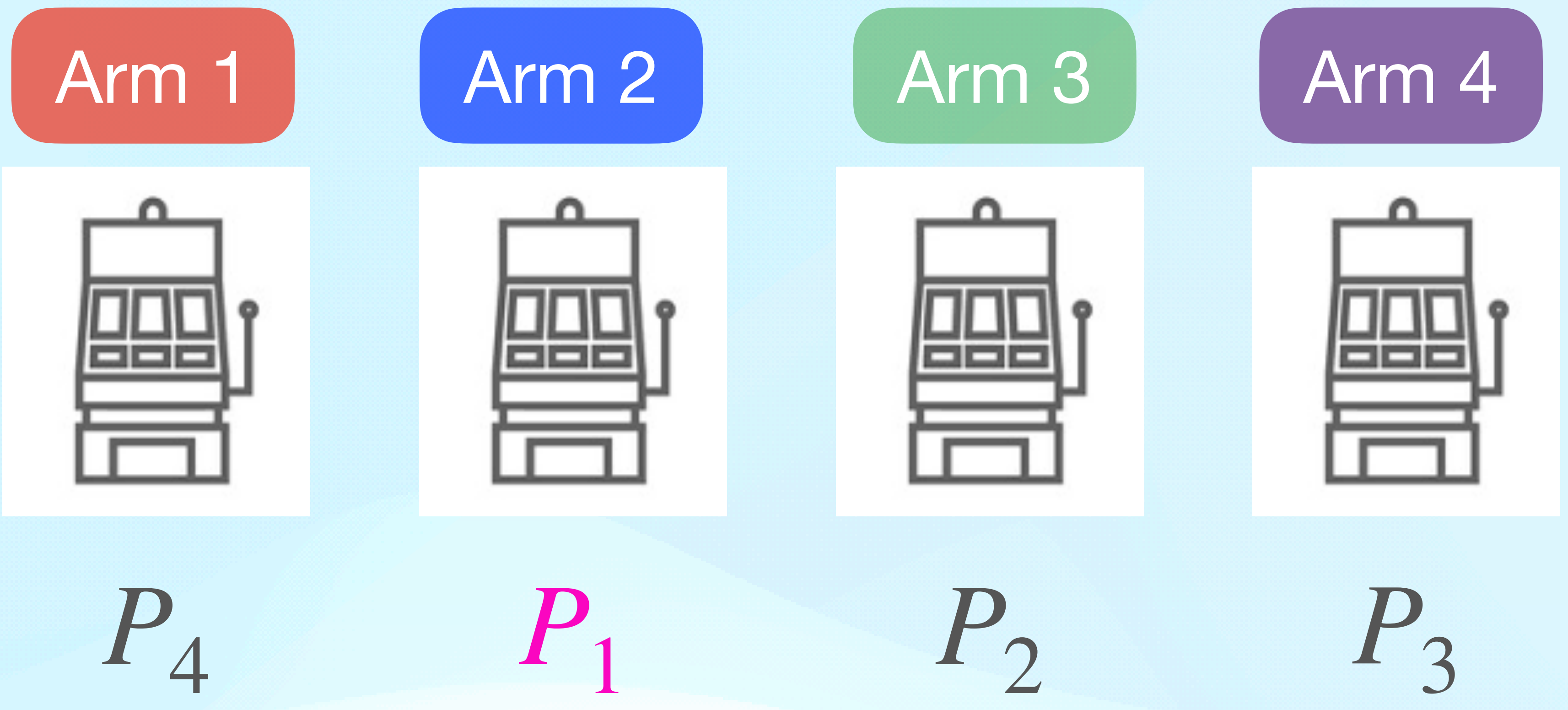
$$\inf_{\pi \in \Pi(\delta)} \mathbb{E}^\pi[\text{stopping time under } \pi] \sim \Theta(\log(1/\delta))$$

goal To determine the best arm quickly and with high confidence

NUS National University of Singapore

* TPM(s): transition probability matrix(ces)

PROBLEM SETUP & OBJECTIVE



- A multi-armed bandit with $K \geq 2$ arms
- Arm: ergodic discrete-time Markov chain on finite state space $\mathcal{S}$
- Arms are restless
- Given TPMs* $P_1, \ldots, P_K$ and a permutation $\sigma : [K] \rightarrow [K]$, the TPM of arm a is $P_{\sigma(a)}$
- TPMs are known, σ is unknown
- Stationary distribution of TPM P_k is μ_k
- Given $f : \mathcal{S} \rightarrow \mathbb{R}$, the best arm

$$a^\star := \arg\max_a \langle f, \mu_{\sigma(a)} \rangle = \arg\max_a \sum_{i \in \mathcal{S}} f(i)\ \mu_{\sigma(a)}(i)$$

accurately $\Pi(\delta) = \{\pi : P^\pi(\hat{a} \neq a^\star) \leq \delta\}$

quickly $\inf_{\pi \in \Pi(\delta)} \mathbb{E}^\pi[\text{stopping time under } \pi]$

$$\inf_{\pi \in \Pi(\delta)} \mathbb{E}^\pi[\text{stopping time under } \pi] \sim \Theta(\log(1/\delta))$$

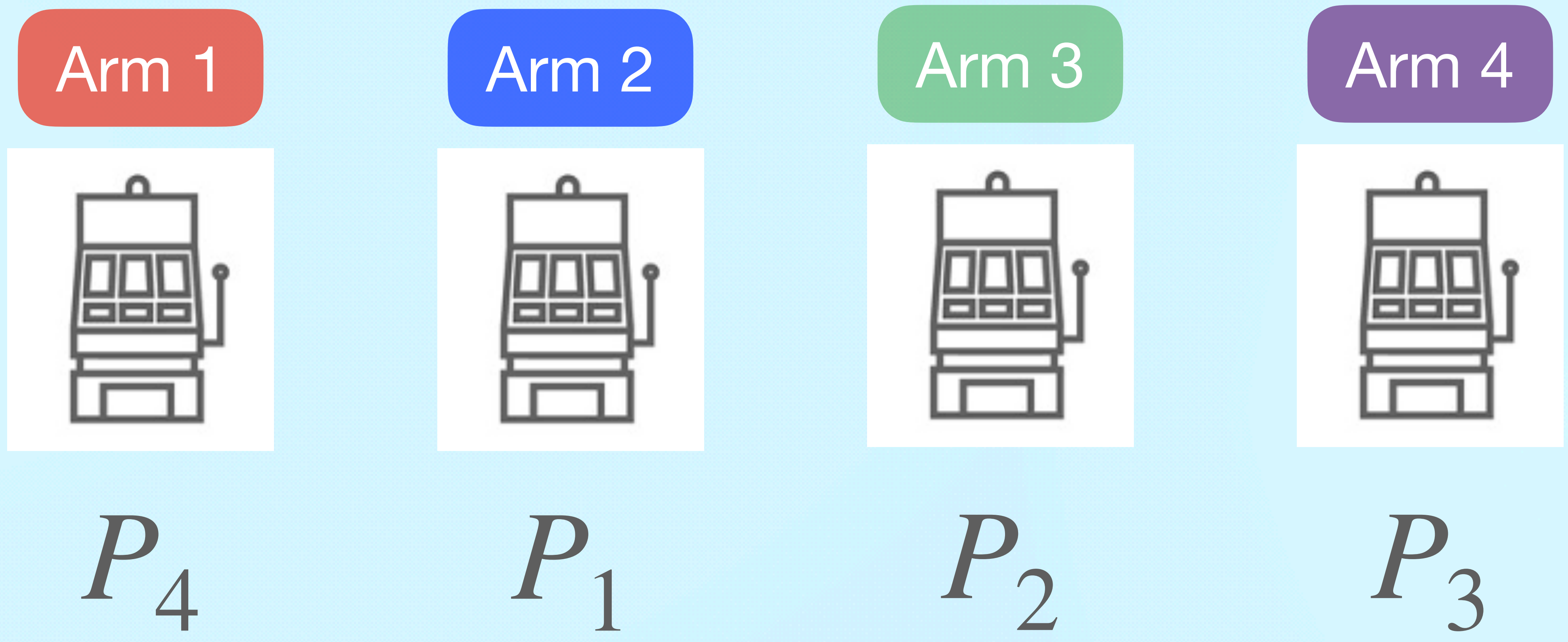
characterise or bound $\boxed{\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)}}$

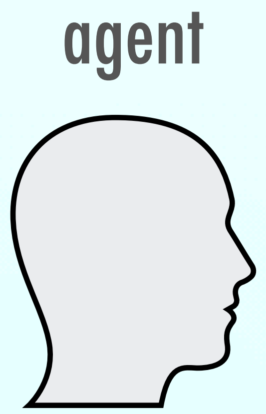
goal **To determine the best arm quickly and with high confidence**

* TPM(s): transition probability matrix(ces)

PRELIMINARIES


Arm 1
Arm 2
Arm 3
Arm 4
P_4
P_1
P_2
P_3
agent


Arm 1
Arm 2
Arm 3
Arm 4
P_4
P_1
P_2
P_3


agent

Arm 1

P_4

Arm 2

P_1

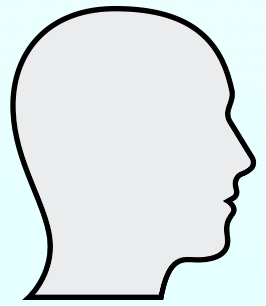
Arm 3

P_2

Arm 4

P_3

agent



t								
a								
$X_{t,a}$								

Arm 1

P_4

Arm 2

P_1

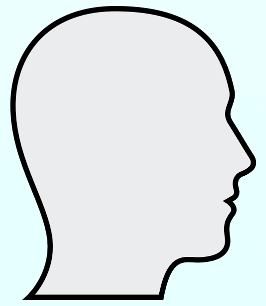
Arm 3

P_2

Arm 4

P_3

agent



t	0							
a	1							
$X_{t,a}$								

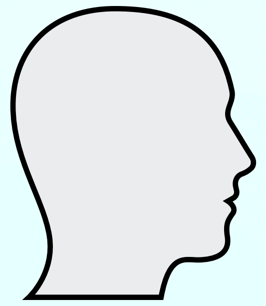
Arm 1 | Arm 2 | Arm 3 | Arm 4

P_4 | P_1 | P_2 | P_3

$t = 0$ $X_{0,1}$

agent



t	0							
a	1							
$X_{t,a}$	$X_{0,1}$							

Arm 1

P_4

Arm 2

P_1

Arm 3

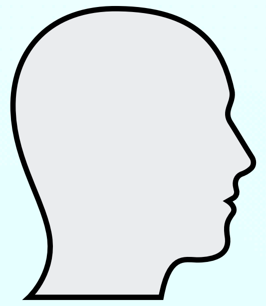
P_2

Arm 4

P_3

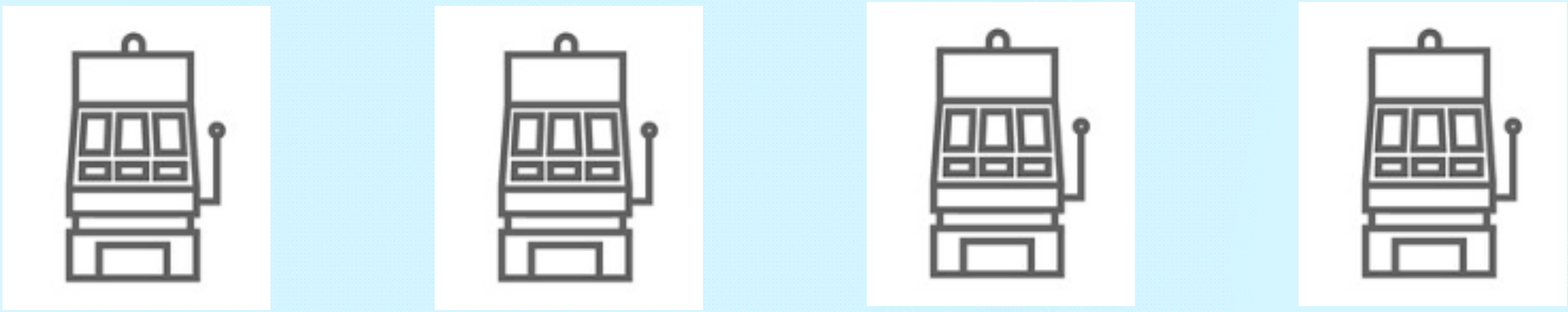
$t = 0$ $X_{0,1}$

agent



t	0	1						
a	1	2						
$X_{t,a}$	$X_{0,1}$							

Arm 1 Arm 2 Arm 3 Arm 4

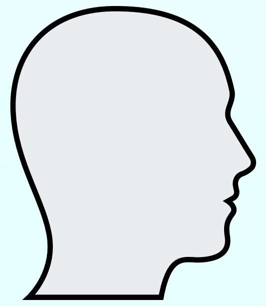


P_4 P_1 P_2 P_3

$t = 0$ $\color{red}{X_{0,1}}$

$t = 1$ $\color{blue}{X_{1,2}}$

agent

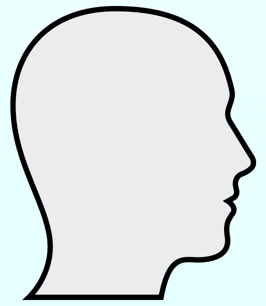


t	0	1						
a	1	2						
$X_{t,a}$	$\color{red}{X_{0,1}}$	$\color{blue}{X_{1,2}}$						

Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

	Arm 1	Arm 2	Arm 3	Arm 4
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		

agent



t	0	1	2					
a	1	2	3					
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$						

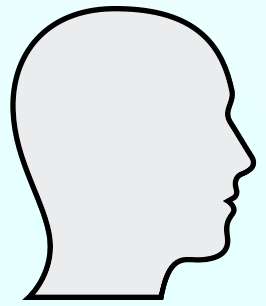
Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

$t = 0$ $X_{0,1}$

$t = 1$ $X_{1,2}$

$t = 2$ $X_{2,3}$

agent



t	0	1	2					
a	1	2	3					
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$					

Arm 1 | Arm 2 | Arm 3 | Arm 4

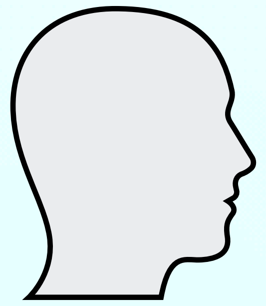
P_4 $\quad P_1$ $\quad P_2$ $\quad P_3$

$t = 0$ $\quad X_{0,1}$

$t = 1$ $\quad X_{1,2}$

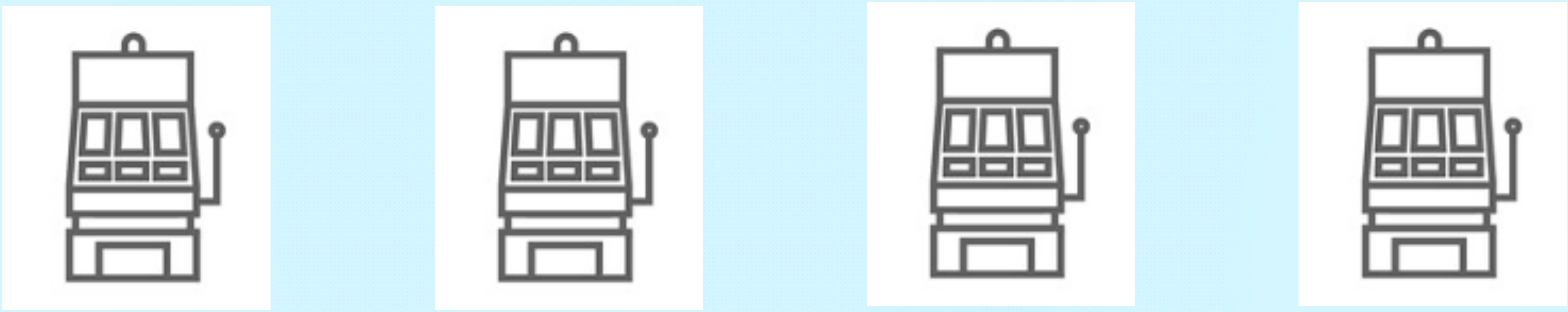
$t = 2$ $\quad X_{2,3}$

agent



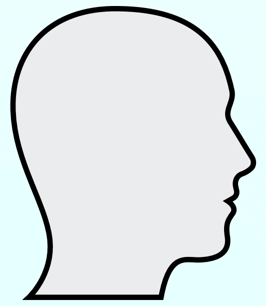
t	0	1	2	3				
a	1	2	3	4				
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$					

Arm 1	Arm 2	Arm 3	Arm 4



	P_4	P_1	P_2	P_3
$t=0$	$X_{0,1}$			
$t=1$		$X_{1,2}$		
$t=2$			$X_{2,3}$	
$t=3$				$X_{3,4}$

agent



t	0	1	2	3				
a	1	2	3	4				
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$				

Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

	Arm 1	Arm 2	Arm 3	Arm 4
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$

agent

t	0	1	2	3	4			
a	1	2	3	4	3			
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$				

Arm 1

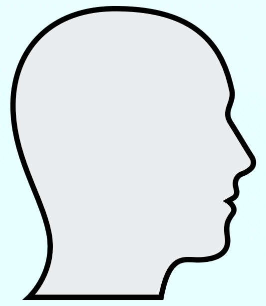
Arm 2

Arm 3

Arm 4


	Arm 1	Arm 2	Arm 3	Arm 4
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	

agent

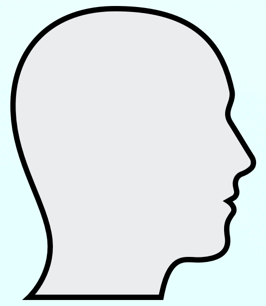


t	0	1	2	3	4			
a	1	2	3	4	3			
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$			

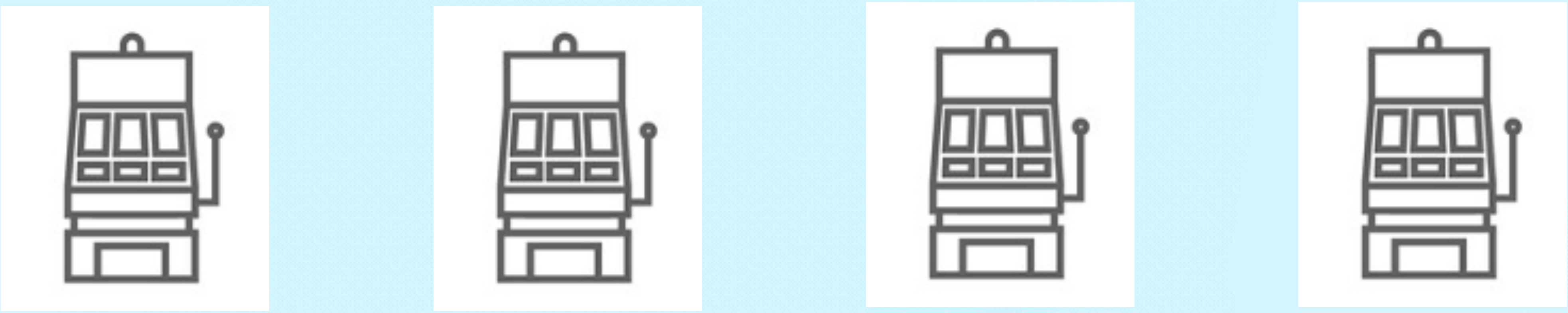
Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

	Arm 1	Arm 2	Arm 3	Arm 4
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	

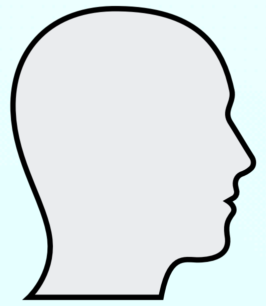
agent



t	0	1	2	3	4	5		
a	1	2	3	4	3	3		
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$			

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	

agent



t	0	1	2	3	4	5	6	
a	1	2	3	4	3	3	2	
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$			

Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

$t = 0$ $X_{0,1}$

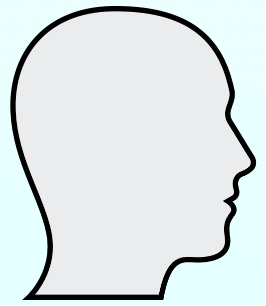
$t = 1$ $X_{1,2}$

$t = 2$ $X_{2,3}$

$t = 3$ $X_{3,4}$

$t = 4$ $X_{4,3}$

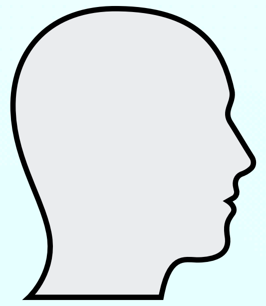
agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$			

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

agent

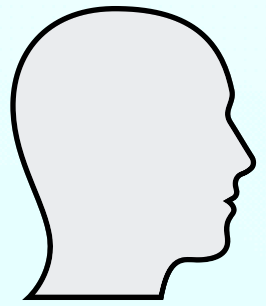


t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

$t = 8$

agent

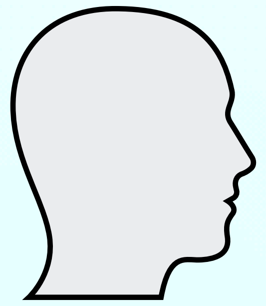


t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

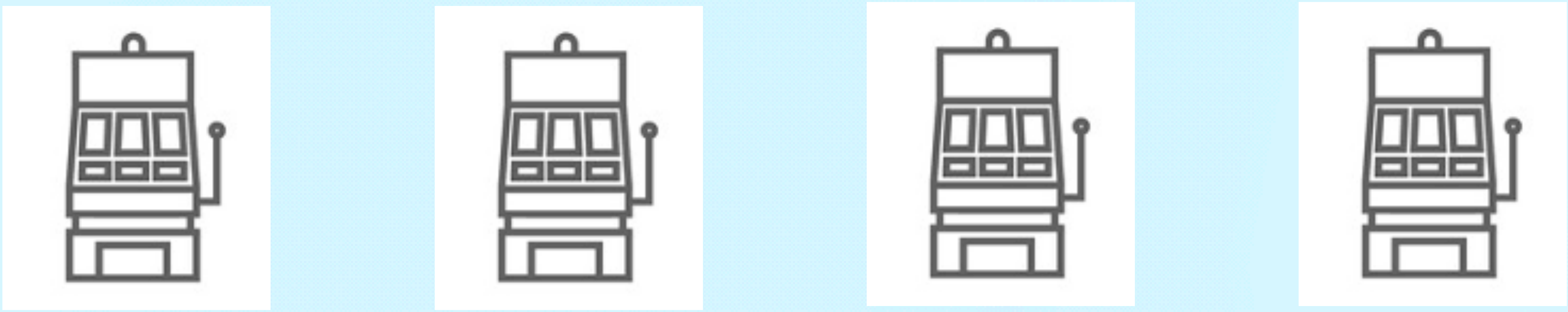
agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$

	Arm 1	Arm 2	Arm 3	Arm 4



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

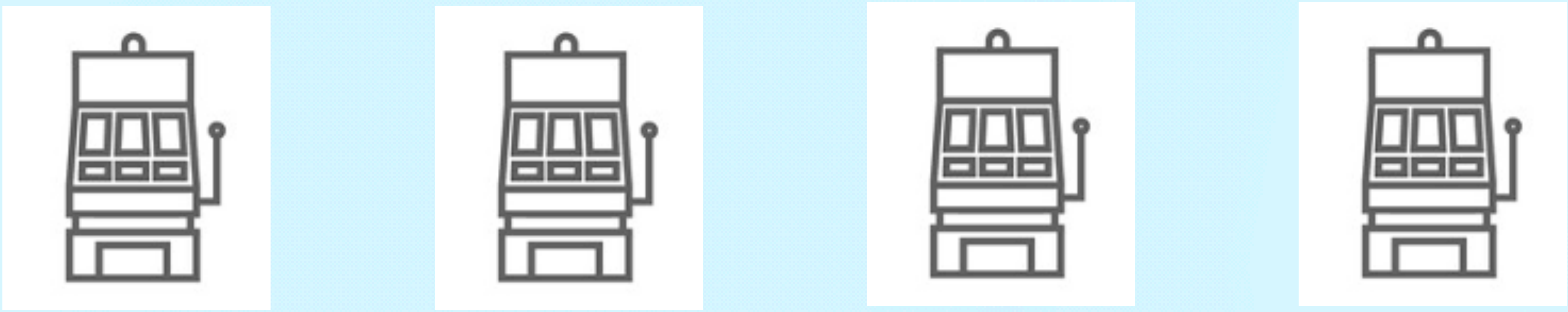
$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

agent

t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

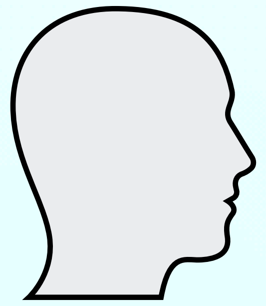
$t = 8 \quad d_1(t) = 1$

	Arm 1	Arm 2	Arm 3	Arm 4
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				



$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

agent



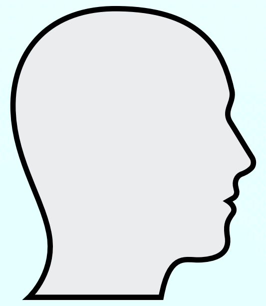
t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2$

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

agent



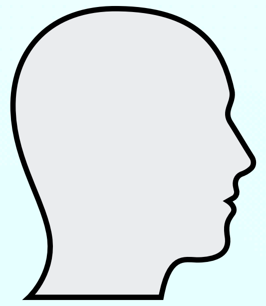
t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3$

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

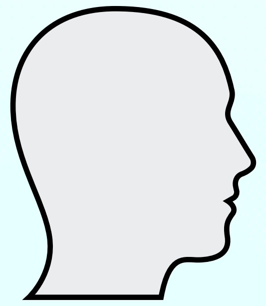
$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

	Arm 1	Arm 2	Arm 3	Arm 4
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

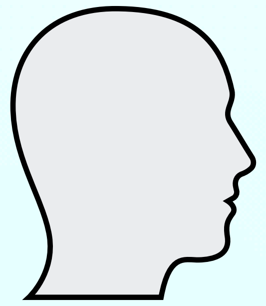
$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$i_1(t) = X_{7,1}$

Arm 1

Arm 2

Arm 3

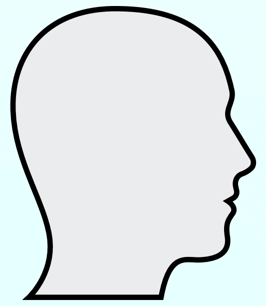
Arm 4

	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent

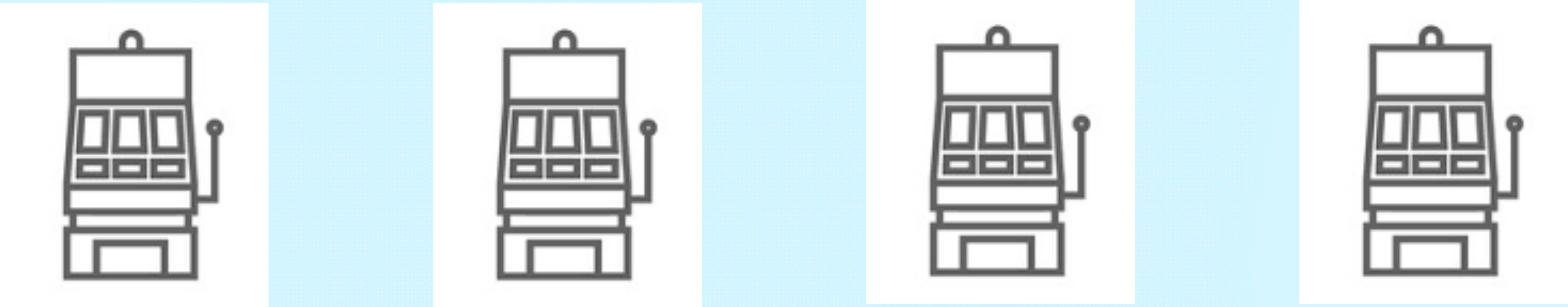


t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$i_1(t) = X_{7,1} \quad i_2(t) = X_{6,2}$

Arm 1	Arm 2	Arm 3	Arm 4

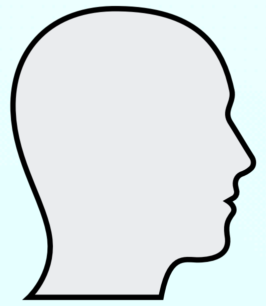


	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$ $\quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$i_1(t) = X_{7,1} \quad i_2(t) = X_{6,2} \quad i_3(t) = X_{5,3}$

Arm 1

Arm 2

Arm 3

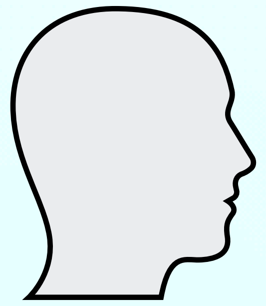
Arm 4


	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$ $\quad d_1(t) = 1$ $\quad d_2(t) = 2$ $\quad d_3(t) = 3$ $\quad d_4(t) = 5$

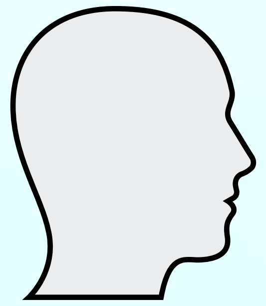
$i_1(t) = X_{7,1}$ $\quad i_2(t) = X_{6,2}$ $\quad i_3(t) = X_{5,3}$ $\quad i_4(t) = X_{3,4}$

	Arm 1	Arm 2	Arm 3	Arm 4
				
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$ $\quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$i_1(t) = X_{7,1} \quad i_2(t) = X_{6,2} \quad i_3(t) = X_{5,3} \quad i_4(t) = X_{3,4}$

Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

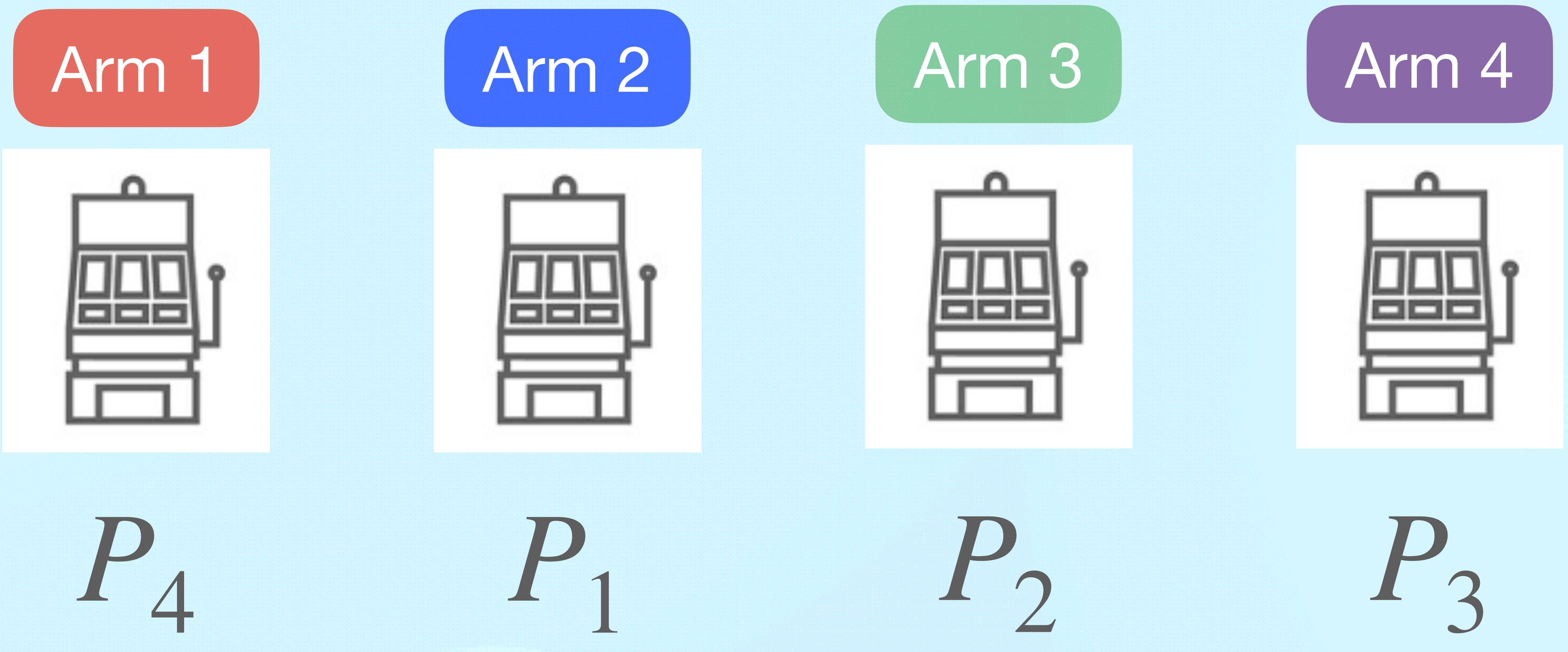
agent

t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$ $\quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$\quad i_1(t) = X_{7,1} \quad i_2(t) = X_{6,2} \quad i_3(t) = X_{5,3} \quad i_4(t) = X_{3,4}$

$t = 5$



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

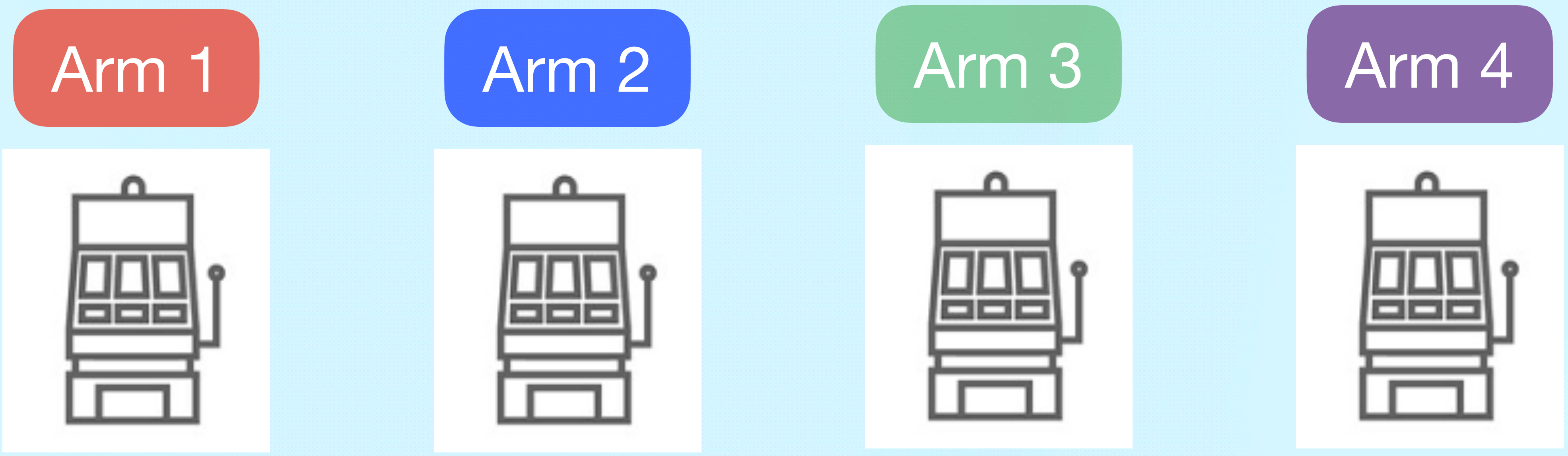
agent

t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$ $d_1(t) = 1$ $d_2(t) = 2$ $d_3(t) = 3$ $d_4(t) = 5$

$i_1(t) = X_{7,1}$ $i_2(t) = X_{6,2}$ $i_3(t) = X_{5,3}$ $i_4(t) = X_{3,4}$

$t = 5$ $d_1(t) = 5$ $d_2(t) = 4$ $d_3(t) = 1$ $d_4(t) = 2$

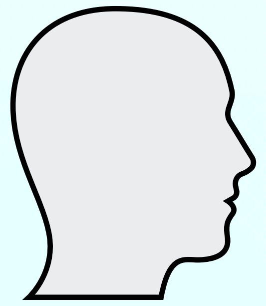


	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent



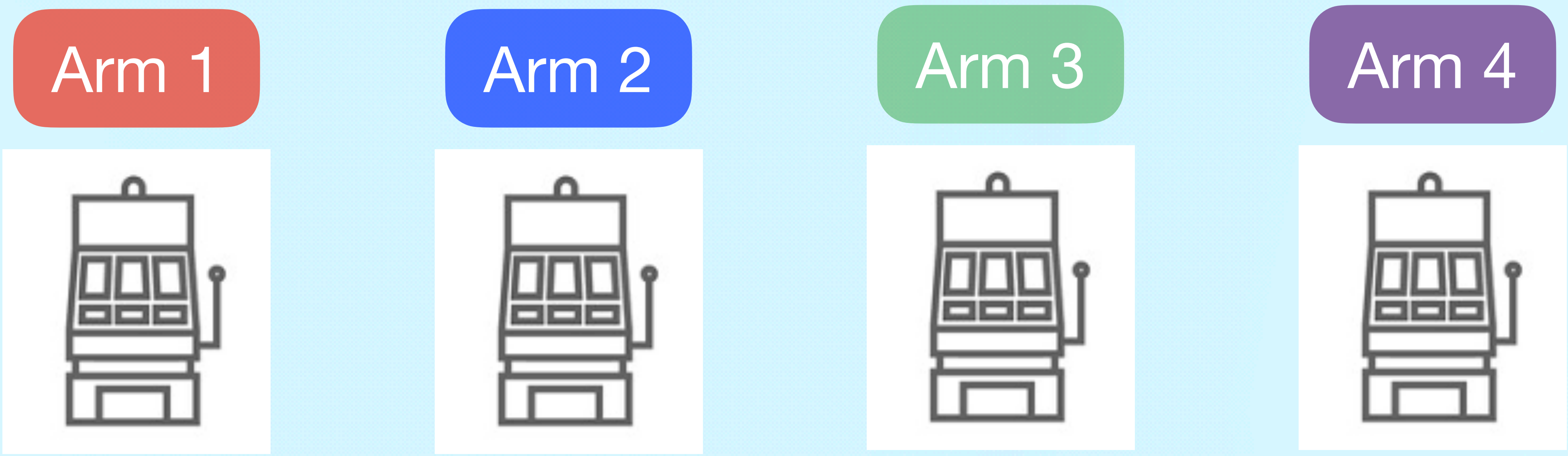
t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8$ $\quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$\quad i_1(t) = X_{7,1} \quad i_2(t) = X_{6,2} \quad i_3(t) = X_{5,3} \quad i_4(t) = X_{3,4}$

$t = 5$ $\quad d_1(t) = 5 \quad d_2(t) = 4 \quad d_3(t) = 1 \quad d_4(t) = 2$

$\quad i_1(t) = X_{0,1} \quad i_2(t) = X_{1,2} \quad i_3(t) = X_{4,3} \quad i_4(t) = X_{3,4}$

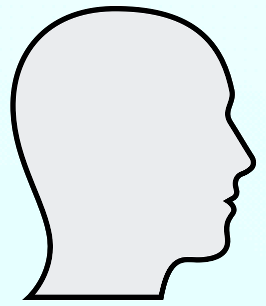


	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent



t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

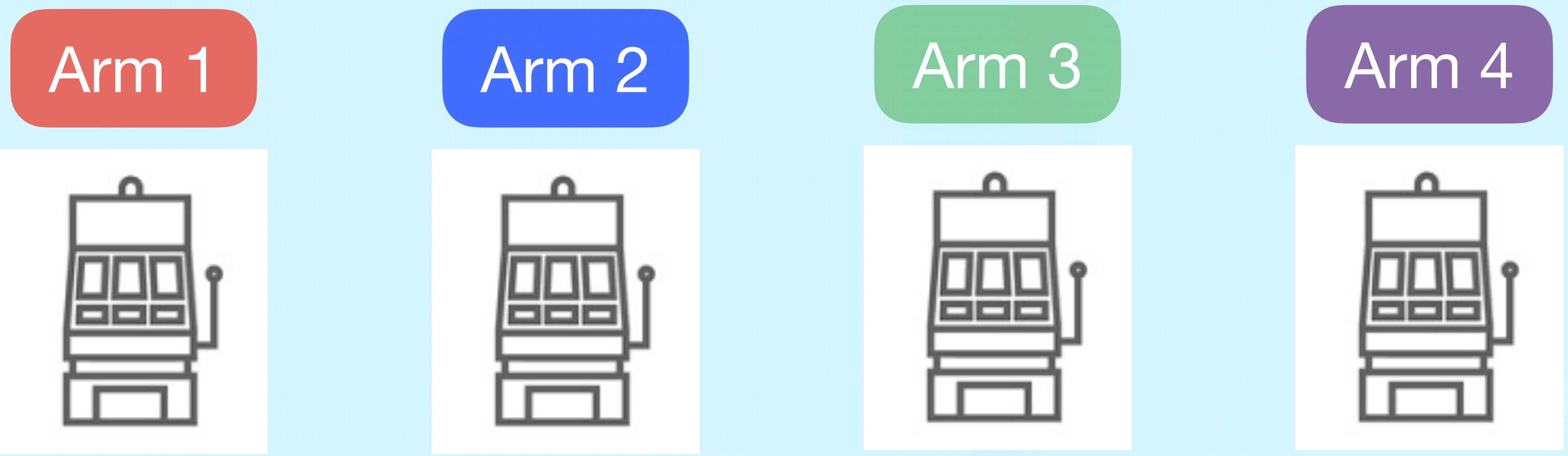
$t = 8$ $d_1(t) = 1$ $d_2(t) = 2$ $d_3(t) = 3$ $d_4(t) = 5$

$i_1(t) = X_{7,1}$ $i_2(t) = X_{6,2}$ $i_3(t) = X_{5,3}$ $i_4(t) = X_{3,4}$

$t = 5$ $d_1(t) = 5$ $d_2(t) = 4$ $d_3(t) = 1$ $d_4(t) = 2$

$i_1(t) = X_{0,1}$ $i_2(t) = X_{1,2}$ $i_3(t) = X_{4,3}$ $i_4(t) = X_{3,4}$

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$$



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent

t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

$t = 8 \quad d_1(t) = 1 \quad d_2(t) = 2 \quad d_3(t) = 3 \quad d_4(t) = 5$

$\quad i_1(t) = X_{7,1} \quad i_2(t) = X_{6,2} \quad i_3(t) = X_{5,3} \quad i_4(t) = X_{3,4}$

$t = 5 \quad d_1(t) = 5 \quad d_2(t) = 4 \quad d_3(t) = 1 \quad d_4(t) = 2$

$\quad i_1(t) = X_{0,1} \quad i_2(t) = X_{1,2} \quad i_3(t) = X_{4,3} \quad i_4(t) = X_{3,4}$

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$$

$$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$$

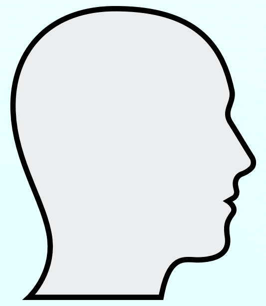
Arm 1	Arm 2	Arm 3	Arm 4
			
P_4	P_1	P_2	P_3

$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			
$t = 8$				

$d_a(t)$: # time instants ago arm a was previously observed w.r.t. time t

$i_a(t)$: last observed state of arm a w.r.t. time t

agent

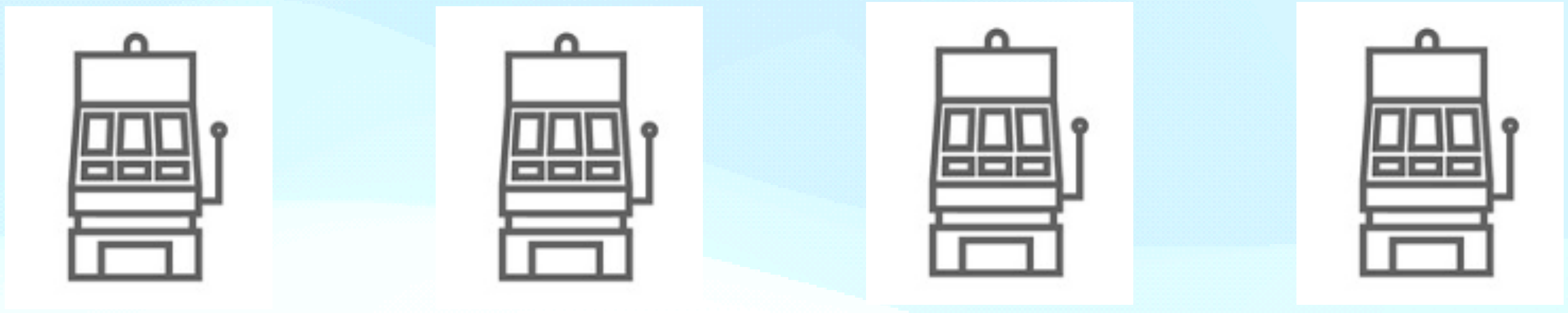


t	0	1	2	3	4	5	6	7
a	1	2	3	4	3	3	2	1
$X_{t,a}$	$X_{0,1}$	$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,3}$	$X_{5,3}$	$X_{6,2}$	$X_{7,1}$

AN MDP

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$$

$$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$$



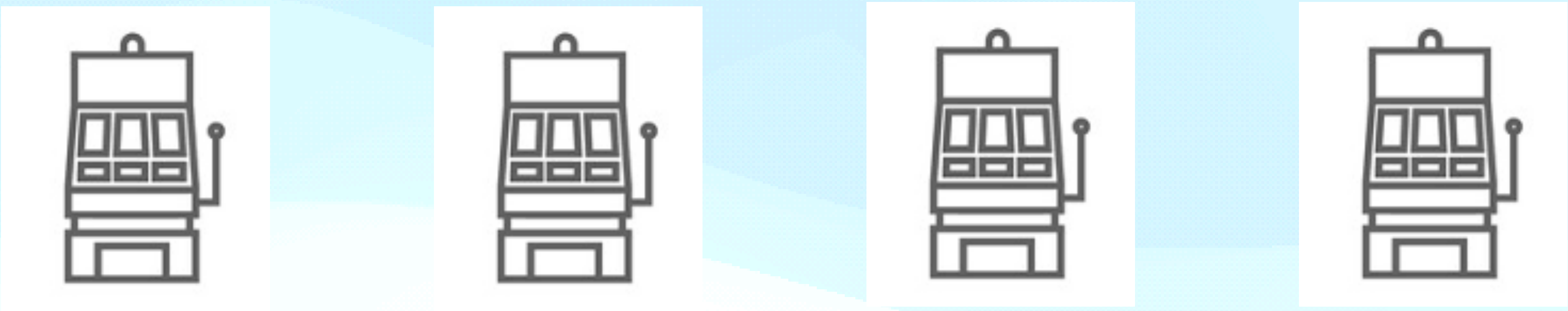
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

AN MDP

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$$

$$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$



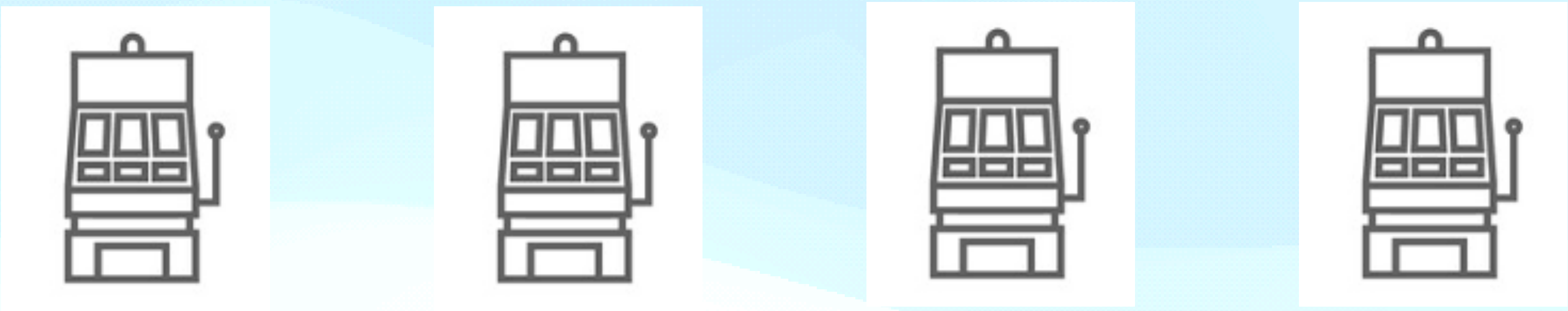
	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

AN MDP

$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$

$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow \underset{\text{arm pulled at time } t}{A_t} \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

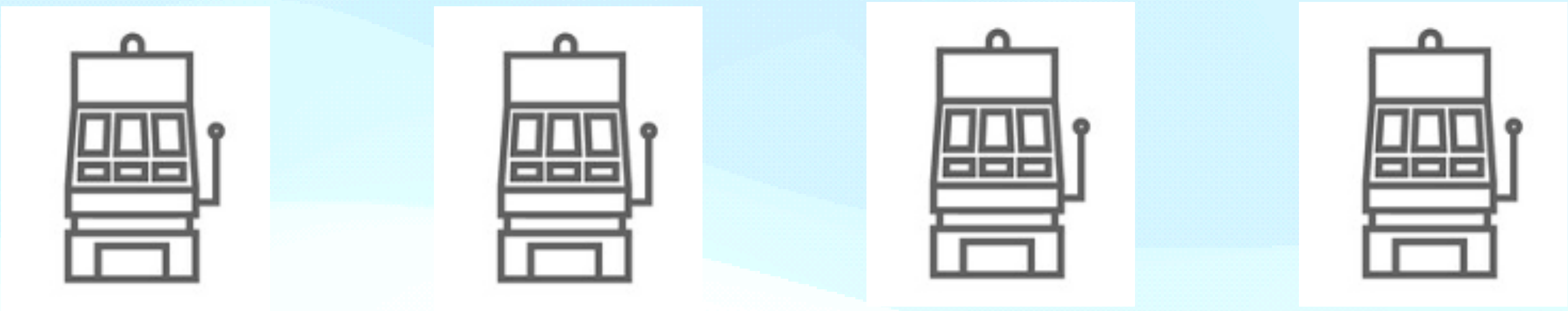
NUS National University of Singapore

AN MDP

$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$

$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow \underbrace{A_t}_{\text{arm pulled at time } t} \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

MDP

AN MDP

$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$

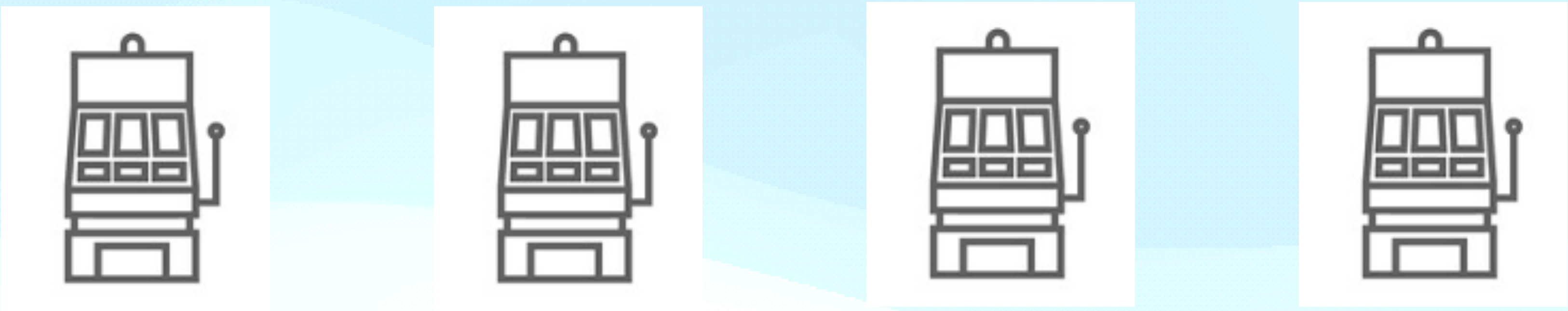
$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$

arm pulled at time t

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$

MDP



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

NUS National University of Singapore

AN MDP

$$\underline{d}(t) = (d_1(t), \dots, d_K(t))$$

$$\underline{i}(t) = (i_1(t), \dots, i_K(t))$$

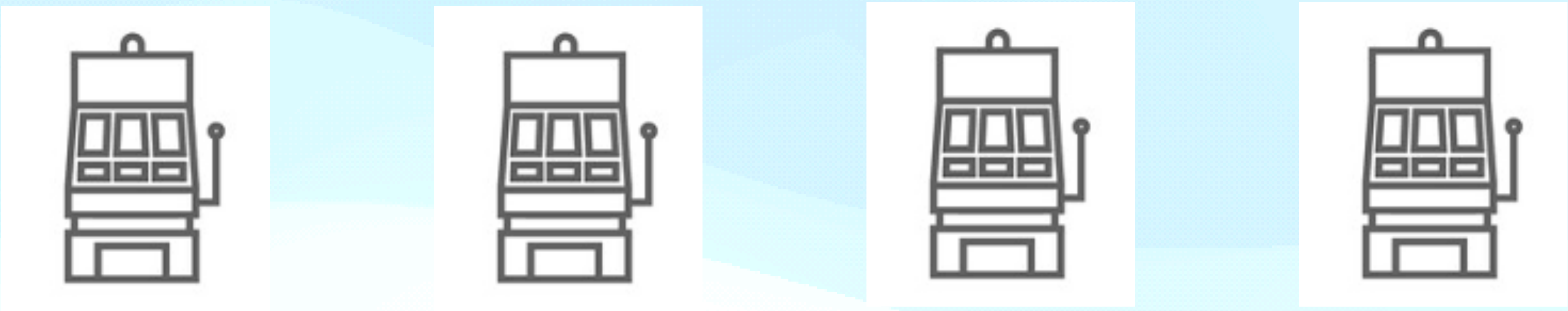
$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$

arm pulled at time t

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$

action space $\{1, \dots, K\}$

MDP



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

NUS National University of Singapore

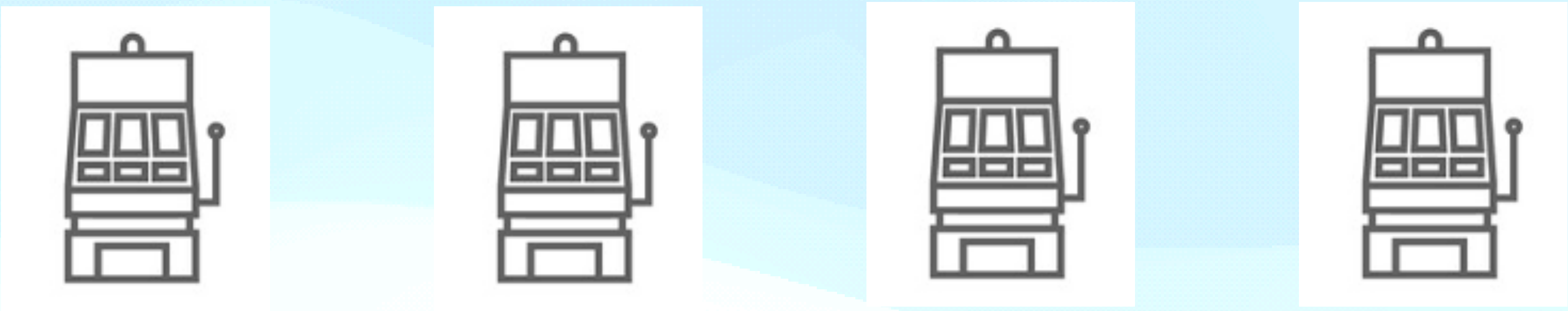
AN MDP

$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$

$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$

arm pulled at time t



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

MDP

state space: $\mathbb{S} = \{(\underline{d}, \underline{i})\}$

action space: $\{1, \ldots, K\}$

state at time t: $(\underline{d}(t), \underline{i}(t))$

AN MDP

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$$

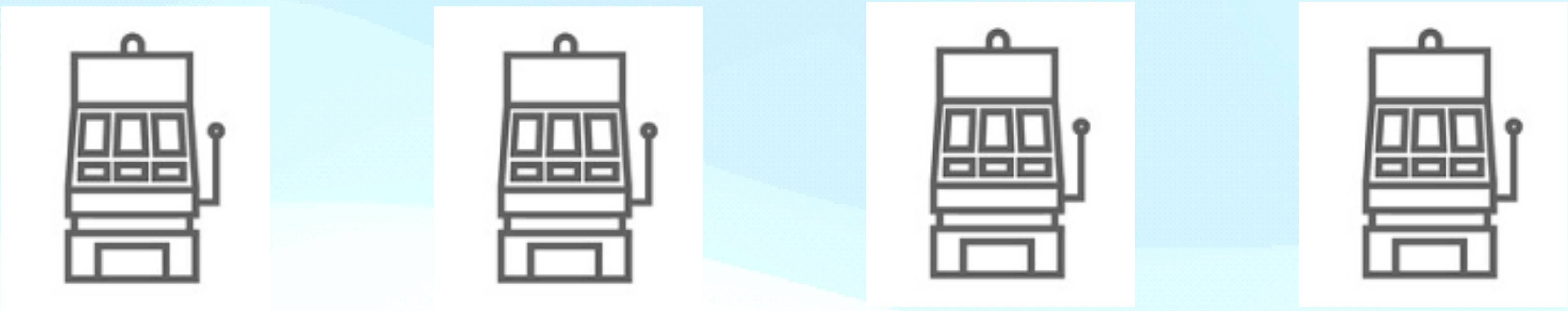
$$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$

arm pulled at time t

MDP

state space	$\mathbb{S} = \{(\underline{d}, \underline{i})\}$
action space	$\{1, \ldots, K\}$
state at time t	$(\underline{d}(t), \underline{i}(t))$
action at time t	A_t



	P_4	P_1	P_2	P_3
$t=0$	$X_{0,1}$			
$t=1$		$X_{1,2}$		
$t=2$			$X_{2,3}$	
$t=3$				$X_{3,4}$
$t=4$			$X_{4,3}$	
$t=5$			$X_{5,3}$	
$t=6$		$X_{6,2}$		
$t=7$	$X_{7,1}$			

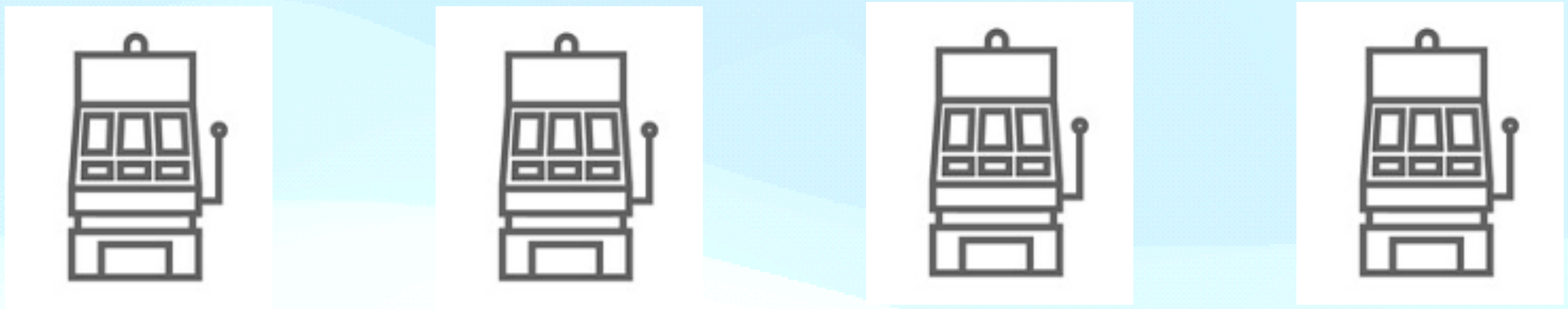
AN MDP

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$$

$$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$

arm pulled at time t



	P_4	P_1	P_2	P_3
$t=0$	$X_{0,1}$			
$t=1$		$X_{1,2}$		
$t=2$			$X_{2,3}$	
$t=3$				$X_{3,4}$
$t=4$			$X_{4,3}$	
$t=5$			$X_{5,3}$	
$t=6$		$X_{6,2}$		
$t=7$	$X_{7,1}$			

MDP

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$

action space $\{1, \ldots, K\}$

state at time t $(\underline{d}(t), \underline{i}(t))$

action at time t A_t

transition probabilities $\mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) = (P_{\sigma(a)})^{d_a}(i'_a \mid i_a)$

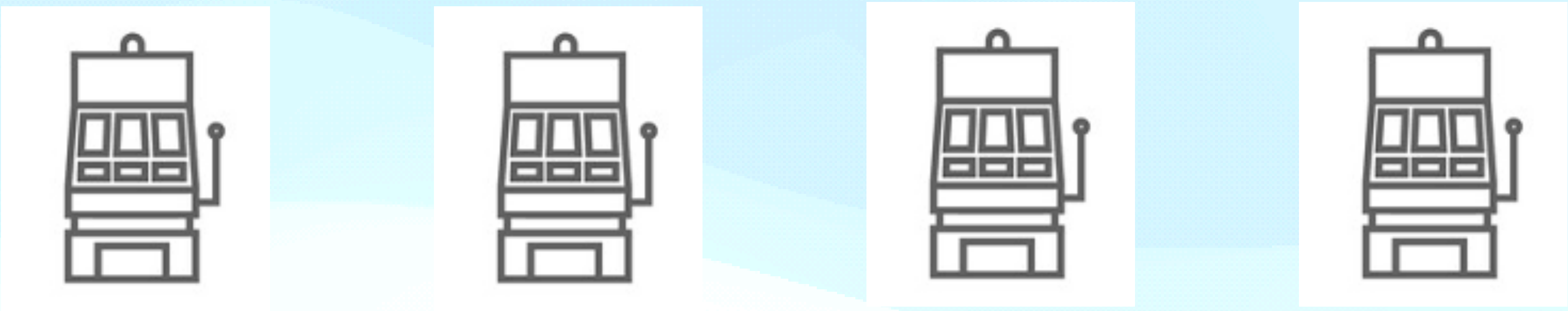
AN MDP

$\underline{d}(t) = (d_1(t), \ldots, d_K(t))$

$\underline{i}(t) = (i_1(t), \ldots, i_K(t))$

$$(\underline{d}(t), \underline{i}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{i}(t+1)) \longrightarrow A_{t+1} \longrightarrow \cdots\cdots$$

arm pulled at time t



	P_4	P_1	P_2	P_3
$t = 0$	$X_{0,1}$			
$t = 1$		$X_{1,2}$		
$t = 2$			$X_{2,3}$	
$t = 3$				$X_{3,4}$
$t = 4$			$X_{4,3}$	
$t = 5$			$X_{5,3}$	
$t = 6$		$X_{6,2}$		
$t = 7$	$X_{7,1}$			

MDP

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$

action space $\{1, \ldots, K\}$

state at time t $(\underline{d}(t), \underline{i}(t))$

action at time t A_t

transition probabilities $\mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) = (P_{\sigma(a)})^{d_a}(i'_a \mid i_a)$

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)}$$

NUS National University of Singapore

CONVERSE: LOWER BOUND

LOWER BOUND

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)}$$

$$C = (P_{\sigma(1)}, \ldots, P_{\sigma(K)})$$

LOWER BOUND

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)}$$

$$C = (P_{\sigma(1)}, \ldots, P_{\sigma(K)})$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^\star(C)}$$

LOWER BOUND

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)}$$

$$C = (P_{\sigma(1)}, \ldots, P_{\sigma(K)})$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^{\star}(C)}$$

$$T^{\star}(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

NUS National University of Singapore

LOWER BOUND

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)}$$

$$C = (P_{\sigma(1)}, \ldots, P_{\sigma(K)})$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^{\star}(C)}$$

$$T^{\star}(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) \, k_{CC'}(\underline{d}, \underline{i}, a)$$

LOWER BOUND

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)}$$

$$C = (P_{\sigma(1)}, \ldots, P_{\sigma(K)})$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^{\star}(C)}$$

$$T^{\star}(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) k_{CC'}(\underline{d}, \underline{i}, a)$$

$$\nu(\underline{d}, \underline{i}, a) \geq 0 \text{ for all } (\underline{d}, \underline{i}, a)$$

$$\sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) = 1$$

NUS National University of Singapore

LOWER BOUND

characterise or bound

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)}$$

$$C = (P_{\sigma(1)}, \dots, P_{\sigma(K)})$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^{\star}(C)}$$

$$T^{\star}(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a) k_{CC'}(\underline{d},\underline{i},a)$$

best arm = 1	1,2,3,4	1,2,4,3	1,3,2,4
	1,3,4,2	1,4,2,3	1,4,3,2
best arm = 2	2,1,3,4	2,1,4,3	3,1,2,4
	3,1,4,2	4,1,2,3	4,1,3,2
best arm = 3	2,3,1,4	2,4,1,3	3,2,1,4
	3,4,1,2	4,2,1,3	4,3,1,2
best arm = 4	2,3,4,1	2,4,3,1	3,2,4,1
	3,4,2,1	4,2,3,1	4,3,2,1

C

$Alt(C)$

$\nu(\underline{d},\underline{i},a) \geq 0$ for all $(\underline{d},\underline{i},a)$

$$\sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a) = 1$$

NUS National University of Singapore

SOME OBSERVATIONS - 1

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\; k_{CC'}(\underline{d}, \underline{i}, a)$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^\star(C)}$$

SOME OBSERVATIONS - 1

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\ k_{CC'}(\underline{d},\underline{i},a)$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^\star(C)}$$

$$\min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu_\epsilon(\underline{d},\underline{i},a)\ k_{CC'}(\underline{d},\underline{i},a) \geq T^\star(C) - \epsilon$$

SOME OBSERVATIONS - 1

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\; k_{CC'}(\underline{d}, \underline{i}, a)$$

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^\star(C)}$$

$$\min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu_\epsilon(\underline{d}, \underline{i}, a)\; k_{CC'}(\underline{d}, \underline{i}, a) \geq T^\star(C) - \epsilon$$

$$\frac{\#\text{ times } (\underline{d}, \underline{i}, a) \text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\, (\underline{d}, \underline{i}, a)$$

SOME OBSERVATIONS - 1

$$\frac{\#\text{ times } (\underline{d}, \underline{i}, a) \text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\ (\underline{d}, \underline{i}, a)$$

SOME OBSERVATIONS - 1

$$\frac{\#\text{ times }(\underline{d}, \underline{i}, a)\text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\, (\underline{d}, \underline{i}, a)$$

On the Empirical State-Action Frequencies in Markov Decision Processes Under General Policies

Shie Mannor
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Québec, Canada H3A 2A7, shie@ece.mcgill.ca, www.ece.mcgill.ca/~shie/

John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu, web.mit.edu/~jnt/www/home.html

SOME OBSERVATIONS - 1

$$\frac{\#\text{ times } (\underline{d}, \underline{i}, a) \text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\, (\underline{d}, \underline{i}, a)$$

$$\left\{ \sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\ \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S} \right\}$$

On the Empirical State-Action Frequencies in Markov Decision Processes Under General Policies

Shie Mannor
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Québec, Canada H3A 2A7, shie@ece.mcgill.ca, www.ece.mcgill.ca/~shie/

John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu, web.mit.edu/~jnt/www/home.html

SOME OBSERVATIONS - 1

$$\frac{\#\text{ times } (\underline{d}, \underline{i}, a) \text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\, (\underline{d}, \underline{i}, a)$$

$$\left\{ \sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S} \right\}$$

$$(\underline{d}', \underline{i}')$$

On the Empirical State-Action Frequencies in Markov Decision Processes Under General Policies

Shie Mannor
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Québec, Canada H3A 2A7, shie@ece.mcgill.ca, www.ece.mcgill.ca/˜shie/

John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu, web.mit.edu/˜jnt/www/home.html

SOME OBSERVATIONS - 1

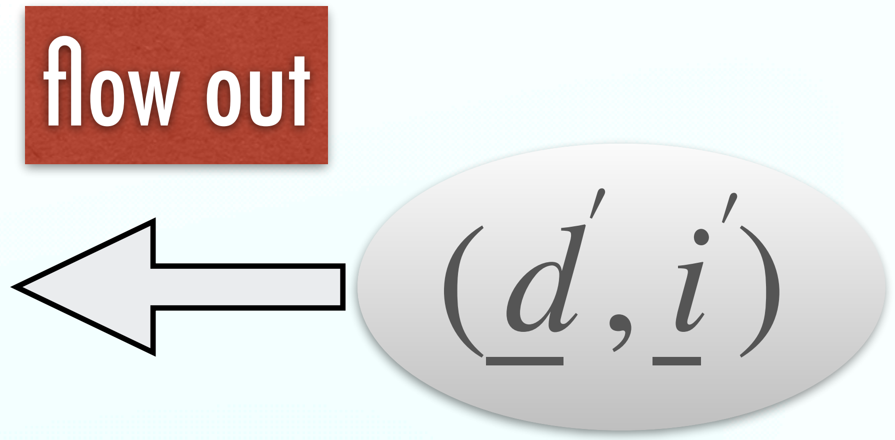
$$\frac{\#\text{ times } (\underline{d}, \underline{i}, a) \text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\ (\underline{d}, \underline{i}, a)$$

$$\left\{ \sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\ \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S} \right\}$$

On the Empirical State-Action Frequencies in Markov Decision Processes Under General Policies

Shie Mannor
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Québec, Canada H3A 2A7, shie@ece.mcgill.ca, www.ece.mcgill.ca/~shie/

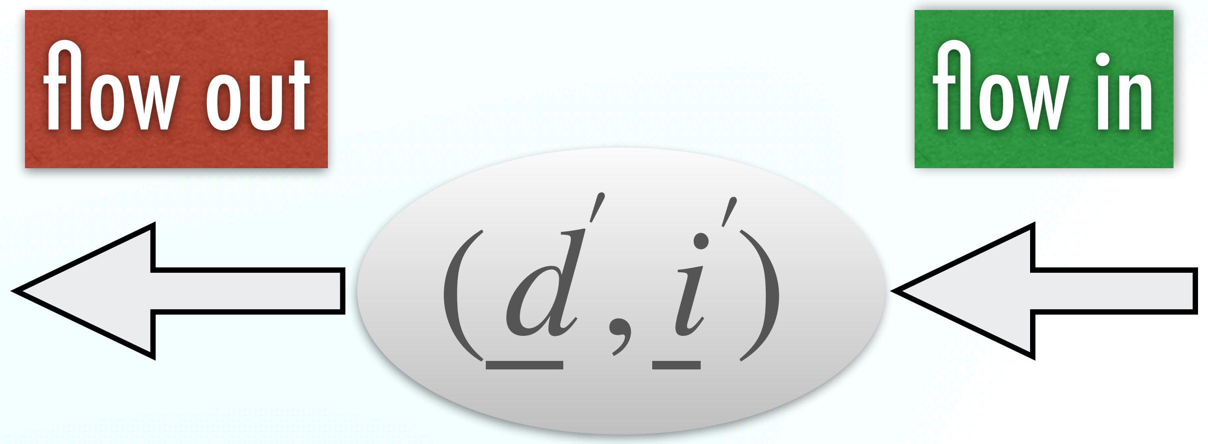
John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu, web.mit.edu/~jnt/www/home.html



SOME OBSERVATIONS - 1

$$\frac{\#\text{ times }(\underline{d}, \underline{i}, a)\text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\, (\underline{d}, \underline{i}, a)$$

$$\left\{ \sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\ \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S} \right\}$$



On the Empirical State-Action Frequencies in Markov Decision Processes Under General Policies

Shie Mannor
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Québec, Canada H3A 2A7, shie@ece.mcgill.ca, www.ece.mcgill.ca/~shie/

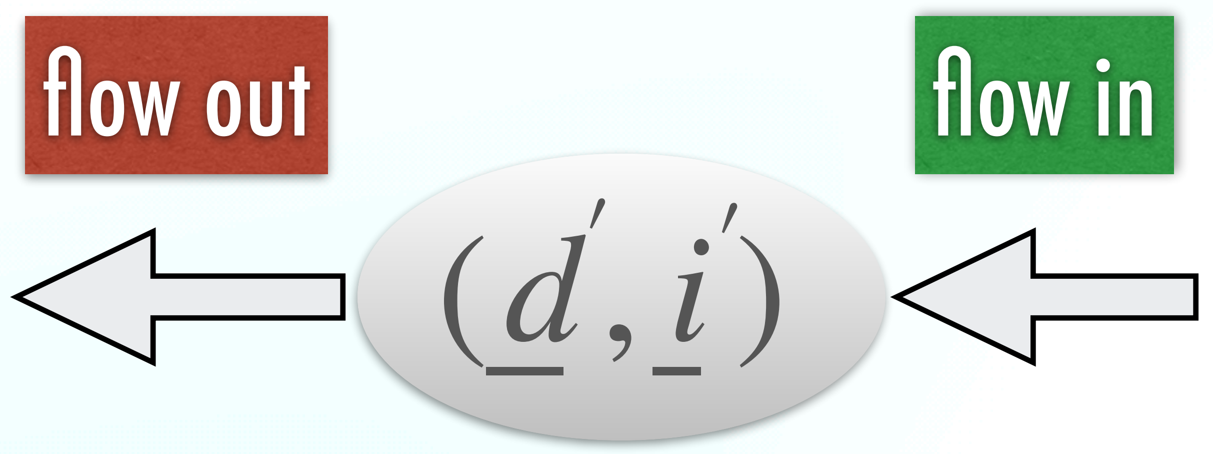
John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu, web.mit.edu/~jnt/www/home.html

SOME OBSERVATIONS - 1

$$\frac{\#\text{ times } (\underline{d}, \underline{i}, a) \text{ is observed up to time } n}{n} \longrightarrow \nu_\epsilon(\underline{d}, \underline{i}, a) \quad \text{as} \quad n \to \infty \quad \forall\, (\underline{d}, \underline{i}, a)$$

flow constraint

$$\left\{ \sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S} \right\}$$



On the Empirical State-Action Frequencies in Markov Decision Processes Under General Policies

Shie Mannor
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Québec, Canada H3A 2A7, shie@ece.mcgill.ca, www.ece.mcgill.ca/~shie/

John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu, web.mit.edu/~jnt/www/home.html

SOME OBSERVATIONS - 2

MDP

state space	$\mathbb{S} = \{(\underline{d}, \underline{i})\}$
action space	$\{1, \ldots, K\}$
state at time t	$(\underline{d}(t), \underline{i}(t))$
action at time t	A_t

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

National University of Singapore

SOME OBSERVATIONS - 2

MDP

state space	$\mathbb{S} = \{(\underline{d}, \underline{i})\}$	countably infinite
action space	$\{1, \ldots, K\}$	
state at time t	$(\underline{d}(t), \underline{i}(t))$	
action at time t	A_t	

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

NUS National University of Singapore

SOME OBSERVATIONS - 2

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$ countably infinite

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

action space $\{1, \ldots, K\}$

MDP

state at time t $(\underline{d}(t), \underline{i}(t))$

action at time t A_t

$$d_a(t) \leq R \quad \forall t, a$$

SOME OBSERVATIONS - 2

MDP

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$ countably infinite

action space $\{1, \ldots, K\}$

state at time t $(\underline{d}(t), \underline{i}(t))$

action at time t A_t

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

$$d_a(t) \leq R \quad \forall t, a$$

$$\left\{ \nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \text{for all } (\underline{d}, \underline{i}) : d_a = R \right\}$$

NUS National University of Singapore

SOME OBSERVATIONS - 2

MDP

state space $\mathbb{S} = \{(\underline{d}, \underline{i})\}$ countably infinite

$$T^\star(C) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\ k_{CC'}(\underline{d}, \underline{i}, a)$$

action space $\{1, \ldots, K\}$

state at time t $(\underline{d}(t), \underline{i}(t))$

action at time t A_t

$$d_a(t) \leq R \quad \forall t, a$$

R-max delay constraint

$$\left\{ \nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \text{for all } (\underline{d}, \underline{i}) : d_a = R \right\}$$

National University of Singapore

MODIFIED OPTIMISATION

$$T_R^\star(P_{\sigma(1)}, \ldots, P_{\sigma(K)}) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a),$$

subject to

$$\sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S}_R,$$

$$\sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) = 1,$$

$$\nu(\underline{d}, \underline{i}, a) \geq 0 \quad \text{for all } (\underline{d}, \underline{i}, a) \in \mathbb{S}_R \times \mathscr{A},$$

$$\nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \text{for all } (\underline{d}, \underline{i}) \in \mathbb{S}_{R,a}, \quad a \in \mathscr{A}.$$

NUS National University of Singapore

MODIFIED OPTIMISATION

$$T_R^\star(P_{\sigma(1)}, \dots, P_{\sigma(K)}) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a),$$

subject to

$$\sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S}_R,$$

$$\sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) = 1,$$

$$\nu(\underline{d}, \underline{i}, a) \geq 0 \quad \text{for all } (\underline{d}, \underline{i}, a) \in \mathbb{S}_R \times \mathscr{A},$$

$$\nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \text{for all } (\underline{d}, \underline{i}) \in \mathbb{S}_{R,a}, \quad a \in \mathscr{A}.$$

MODIFIED OPTIMISATION

$$T_R^\star(P_{\sigma(1)}, \ldots, P_{\sigma(K)}) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a),$$

subject to

$$\sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S}_R,$$

$$\sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) = 1,$$

$$\nu(\underline{d}, \underline{i}, a) \geq 0 \quad \text{for all } (\underline{d}, \underline{i}, a) \in \mathbb{S}_R \times \mathscr{A},$$

$$\nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \text{for all } (\underline{d}, \underline{i}) \in \mathbb{S}_{R,a}, \quad a \in \mathscr{A}.$$

National University of Singapore

MODIFIED OPTIMISATION

$$T_R^\star(P_{\sigma(1)}, \ldots, P_{\sigma(K)}) = \sup_{\nu} \min_{C' \in Alt(C)} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a),$$

subject to

$$\sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, \mathbb{P}((\underline{d}, \underline{i}, a) \to (\underline{d}', \underline{i}')) \quad \text{for all } (\underline{d}', \underline{i}') \in \mathbb{S}_R,$$

$$\sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a) = 1,$$

$$\nu(\underline{d}, \underline{i}, a) \geq 0 \quad \text{for all } (\underline{d}, \underline{i}, a) \in \mathbb{S}_R \times \mathscr{A},$$

$$\nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \text{for all } (\underline{d}, \underline{i}) \in \mathbb{S}_{R,a}, \quad a \in \mathscr{A}.$$

$\boxed{\nu_R^\star = \{\nu_R^\star(\underline{d}, \underline{i}, a)\}}$

ACHIEVABILITY

ALGORITHM FOR BAI - 1

ALGORITHM FOR BAI - 1

$$\boxed{\lambda_{unif}(a \mid \underline{d}, \underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}}$$

ALGORITHM FOR BAI - 1

$$\boxed{\lambda_{unif}(a \mid \underline{d}, \underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}} \longrightarrow \{(\underline{d}(t), \underline{i}(t))\}_{t \geq K}$$ ergodic Markov chain

ALGORITHM FOR BAI - 1

$$\boxed{\lambda_{unif}(a\,|\,\underline{d},\underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}} \longrightarrow \{(\underline{d}(t),\underline{i}(t))\}_{t\geq K} \text{ ergodic Markov chain} \longrightarrow \mu_{unif} = \{\mu_{unif}(\underline{d},\underline{i})\}$$

ALGORITHM FOR BAI - 1

$$\boxed{\lambda_{unif}(a \mid \underline{d}, \underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}}$$

$\longrightarrow \{(\underline{d}(t), \underline{i}(t))\}_{t \geq K}$ **ergodic Markov chain** $\longrightarrow \mu_{unif} = \{\mu_{unif}(\underline{d}, \underline{i})\}$

$$\boxed{\nu_{unif}(\underline{d}, \underline{i}, a) = \begin{cases} \frac{\mu_{unif}(\underline{d}, \underline{i})}{K}, & d_a < R \quad \forall a, \\ \mu_{unif}(\underline{d}, \underline{i}), & d_a = R\,. \end{cases}}$$

ALGORITHM FOR BAI - 1

$$\boxed{\lambda_{unif}(a \mid \underline{d}, \underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R \,. \end{cases}}$$

$\longrightarrow$ $\{(\underline{d}(t), \underline{i}(t))\}_{t \geq K}$ ergodic Markov chain $\longrightarrow$ $\mu_{unif} = \{\mu_{unif}(\underline{d}, \underline{i})\}$

$\downarrow$

$$\boxed{\nu_{unif}(\underline{d}, \underline{i}, a) = \begin{cases} \frac{\mu_{unif}(\underline{d}, \underline{i})}{K}, & d_a < R \quad \forall a, \\ \mu_{unif}(\underline{d}, \underline{i}), & d_a = R \,. \end{cases}}$$

$\nu_R^\star = \{\nu_R^\star(\underline{d}, \underline{i}, a)\}$

ALGORITHM FOR BAI - 1

$$\lambda_{unif}(a\,|\,\underline{d},\underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}$$

$\longrightarrow \{(\underline{d}(t),\underline{i}(t))\}_{t\geq K}$ ergodic Markov chain $\longrightarrow \mu_{unif} = \{\mu_{unif}(\underline{d},\underline{i})\}$

$$\nu_{unif}(\underline{d},\underline{i},a) = \begin{cases} \frac{\mu_{unif}(\underline{d},\underline{i})}{K}, & d_a < R \quad \forall a, \\ \mu_{unif}(\underline{d},\underline{i}), & d_a = R\,. \end{cases}$$

$$\nu_\eta = \eta\, \nu_{unif} + (1-\eta)\, \nu_R^\star \xleftarrow{\eta}$$

$\nu_R^\star = \{\nu_R^\star(\underline{d},\underline{i},a)\}$

NUS National University of Singapore

ALGORITHM FOR BAI - 1

$$\lambda_{unif}(a \mid \underline{d}, \underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}$$

$\longrightarrow \{(\underline{d}(t), \underline{i}(t))\}_{t \geq K}$ **ergodic Markov chain** $\longrightarrow \mu_{unif} = \{\mu_{unif}(\underline{d}, \underline{i})\}$

$$\nu_{unif}(\underline{d}, \underline{i}, a) = \begin{cases} \frac{\mu_{unif}(\underline{d}, \underline{i})}{K}, & d_a < R \quad \forall a, \\ \mu_{unif}(\underline{d}, \underline{i}), & d_a = R\,. \end{cases}$$

$\xleftarrow{\eta}$

$$\nu_\eta = \eta\, \nu_{unif} + (1-\eta)\, \nu_R^\star$$

$\longleftarrow$

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

$$\nu_R^\star = \{\nu_R^\star(\underline{d}, \underline{i}, a)\}$$

ALGORITHM FOR BAI - 1

$$\boxed{\lambda_{unif}(a\,|\,\underline{d},\underline{i}) = \begin{cases} \frac{1}{K}, & d_a < R \quad \forall a, \\ 1, & d_a = R\,. \end{cases}} \longrightarrow \{(\underline{d}(t),\underline{i}(t))\}_{t\geq K} \text{ ergodic Markov chain} \longrightarrow \mu_{unif} = \{\mu_{unif}(\underline{d},\underline{i})\}$$

$$\boxed{\nu_{unif}(\underline{d},\underline{i},a) = \begin{cases} \frac{\mu_{unif}(\underline{d},\underline{i})}{K}, & d_a < R \quad \forall a, \\ \mu_{unif}(\underline{d},\underline{i}), & d_a = R\,. \end{cases}} \xrightarrow{\eta} \nu_\eta = \eta\, \nu_{unif} + (1-\eta)\, \nu_R^\star \longrightarrow \lambda_\eta(a\,|\,\underline{d},\underline{i}) = \frac{\nu_\eta(\underline{d},\underline{i},a)}{\sum_a \nu_\eta(\underline{d},\underline{i},a)}$$

- $\lambda_\eta(a\,|\,\underline{d},\underline{i}) > 0 \implies \lambda_\eta(a\,|\,\underline{d},\underline{i}) \geq \alpha_\eta > 0$
- λ_η rule makes $\{\underline{d}(t),\underline{i}(t)\}$ an ergodic Markov chain

$$\boxed{\nu_R^\star = \{\nu_R^\star(\underline{d},\underline{i},a)\}}$$

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs

δ, η, R

NUS National University of Singapore

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs $\delta, \quad \eta, \quad R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs

δ, η, R

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

estimate of ground truth

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs

δ, η, R

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

estimate of ground truth

- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs $\delta, \quad \eta, \quad R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

estimate of ground truth

- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and $C' \geq \beta_{\delta,\eta,R}$

$\hat{C}(n)$ is the ground truth

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \,|\, \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs $\delta, \ \eta, \ R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

estimate of ground truth

- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and $C' \geq \beta_{\delta,\eta,R}$
  - STOP

$\hat{C}(n)$ is the ground truth

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs $\delta, \quad \eta, \quad R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

estimate of ground truth

- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and $C' \quad \geq \quad \beta_{\delta,\eta,R}$
  - STOP
  - Declare best arm in $\hat{C}(n)$

$\hat{C}(n)$ is the ground truth

NUS National University of Singapore

ALGORITHM FOR BAI - 2

$$\lambda_\eta(a \mid \underline{d}, \underline{i}) = \frac{\nu_\eta(\underline{d}, \underline{i}, a)}{\sum_a \nu_\eta(\underline{d}, \underline{i}, a)}$$

inputs $\delta, \ \eta, \ R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n

estimate of ground truth

- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP
  - Declare best arm in $\hat{C}(n)$

$\hat{C}(n)$ is the ground truth

- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

PERFORMANCE

PERFORMANCE OF THE ALGORITHM $(\pi^\star)$

inputs δ, η, R

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n — estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and $C' \geq \beta_{\delta,\eta,R}$
  - STOP — $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

PERFORMANCE OF THE ALGORITHM ($\pi^\star$)

inputs $\delta, \ \eta, \ R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n — estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP — $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

- Stops in finite time w.p.1

PERFORMANCE OF THE ALGORITHM ($\pi^\star$)

inputs $\delta, \quad \eta, \quad R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

- Stops in finite time w.p.1
- Achieves error probability $\leq \delta$

PERFORMANCE OF THE ALGORITHM ($\pi^\star$)

inputs δ, η, R

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n — estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP — $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

- Stops in finite time w.p.1
- Achieves error probability $\leq \delta$

$$\frac{1}{T^\star(P_1, \ldots, P_K)} \leq \liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \limsup_{\delta \downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{\eta\ T_{unif}(P_1, \ldots, P_K) + (1-\eta)\ T^\star_R(P_1, \ldots, P_K)}$$

PERFORMANCE OF THE ALGORITHM ($\pi^\star$)

inputs δ, η, R

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n — estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP — $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\,\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

NUS National University of Singapore

PERFORMANCE OF THE ALGORITHM ($\pi^\star$)

inputs $\delta, \quad \eta, \quad R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n — estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP — $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \mid \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

$$\frac{1}{T^\star(P_1, \ldots, P_K)} \leq \liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \limsup_{\substack{\delta \downarrow 0 \\ \eta \downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{T_R^\star(P_1, \ldots, P_K)}$$

NUS National University of Singapore

PERFORMANCE OF THE ALGORITHM ($\pi^\star$)

inputs $\delta, \ \eta, \ R$

- $\hat{C}(n) \in \arg\max_{C} \min_{C' \in Alt(C)}$ LLR between C and C' at time n — estimate of ground truth
- If $\min_{C' \in Alt(\hat{C}(n))}$ LLR between $\hat{C}(n)$ and C' $\geq \beta_{\delta,\eta,R}$
  - STOP — $\hat{C}(n)$ is the ground truth
  - Declare best arm in $\hat{C}(n)$
- Else, pull A_{n+1} according to $\lambda_\eta(\cdot \,|\, \underline{d}(n), \underline{i}(n))$ assuming $\hat{C}(n)$ to be the ground truth

$$\frac{1}{T^\star(P_1, \ldots, P_K)} \leq \liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \limsup_{\substack{\delta \downarrow 0 \\ \eta \downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{T^\star_R(P_1, \ldots, P_K)}$$

RELATION BETWEEN $T^\star$ AND $T_R^\star$

RELATION BETWEEN $T^\star$ AND $T_R^\star$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{T_R^\star(P_1,\ldots,P_K)}$$

NUS National University of Singapore

RELATION BETWEEN $T^\star$ AND $T_R^\star$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{T_R^\star(P_1,\ldots,P_K)}$$

A key monotonicity property

RELATION BETWEEN $T^\star$ AND $T_R^\star$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0}\ \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under }\pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under }\pi^\star]}{\log(1/\delta)} \le \frac{1}{T_R^\star(P_1,\ldots,P_K)}$$

A key monotonicity property

$$T_R^\star(P_1,\ldots,P_K) \le T_{R+1}^\star(P_1,\ldots,P_K) \le T^\star(P_1,\ldots,P_K) \quad \forall\ R$$

RELATION BETWEEN $T^\star$ AND $T_R^\star$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \leq \liminf_{\delta\downarrow 0}\ \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{T_R^\star(P_1,\ldots,P_K)}$$

A key monotonicity property

$$T_R^\star(P_1,\ldots,P_K) \leq T_{R+1}^\star(P_1,\ldots,P_K) \leq T^\star(P_1,\ldots,P_K) \quad \forall\ R$$

$\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)$ exists

RELATION BETWEEN $T^\star$ AND $T^\star_R$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0}\ \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under }\pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under }\pi^\star]}{\log(1/\delta)} \le \frac{1}{T^\star_R(P_1,\ldots,P_K)}$$

A key monotonicity property

$$T^\star_R(P_1,\ldots,P_K) \le T^\star_{R+1}(P_1,\ldots,P_K) \le T^\star(P_1,\ldots,P_K) \quad \forall\ R$$

$\lim_{R\to\infty} T^\star_R(P_1,\ldots,P_K)$ exists

IID obs.

NUS National University of Singapore

RELATION BETWEEN $T^\star$ AND $T_R^\star$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{T_R^\star(P_1,\ldots,P_K)}$$

A key monotonicity property

$$T_R^\star(P_1,\ldots,P_K) \le T_{R+1}^\star(P_1,\ldots,P_K) \le T^\star(P_1,\ldots,P_K) \quad \forall\ R$$

$\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)$ exists

IID obs.

$$\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K) = T^\star(P_1,\ldots,P_K)$$

RELATION BETWEEN $T^\star$ AND $T_R^\star$

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0}\ \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{T_R^\star(P_1,\ldots,P_K)}$$

A key monotonicity property

$$T_R^\star(P_1,\ldots,P_K) \le T_{R+1}^\star(P_1,\ldots,P_K) \le T^\star(P_1,\ldots,P_K) \quad \forall\ R$$

$\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)$ exists

IID obs.

$$\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K) = T^\star(P_1,\ldots,P_K)$$

IID obs.

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0}\ \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \limsup_{\substack{\delta\downarrow 0\\ \eta\downarrow 0\\ R\to\infty}} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)}$$

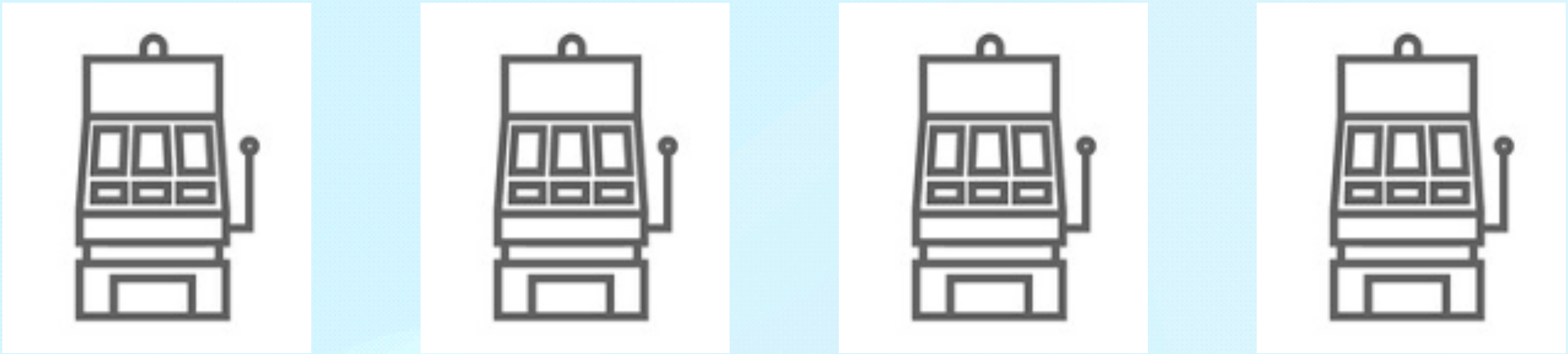
matching bounds

OUR CONTRIBUTIONS

- Converse (lower bound)

$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \geq \frac{1}{T^{\star}(P_1, \ldots, P_K)}$$

- Precise characterisation of $T^{\star}(P_1, \ldots, P_K)$
- Achievability: an algorithm for finding the best arm ($\pi^{\star}$)
- Features of $\pi^{\star}$:
  - $\pi^{\star} \in \Pi(\delta)$ for any $\delta > 0$
  - Practically implementable
- Parameterised upper bound on the expected stopping time of $\pi^{\star}$
- $T_R^{\star}(P_1, \ldots, P_K)$ is monotone increasing in the parameter R



P_4 P_1 P_2 P_3

R-upper bound

$$\limsup_{\substack{\delta \downarrow 0 \\ \eta \downarrow 0}} \frac{E^{\pi^{\star}}[\text{stopping time under } \pi^{\star}]}{\log(1/\delta)} \leq \frac{1}{T_R^{\star}(P_1, \ldots, P_K)}$$

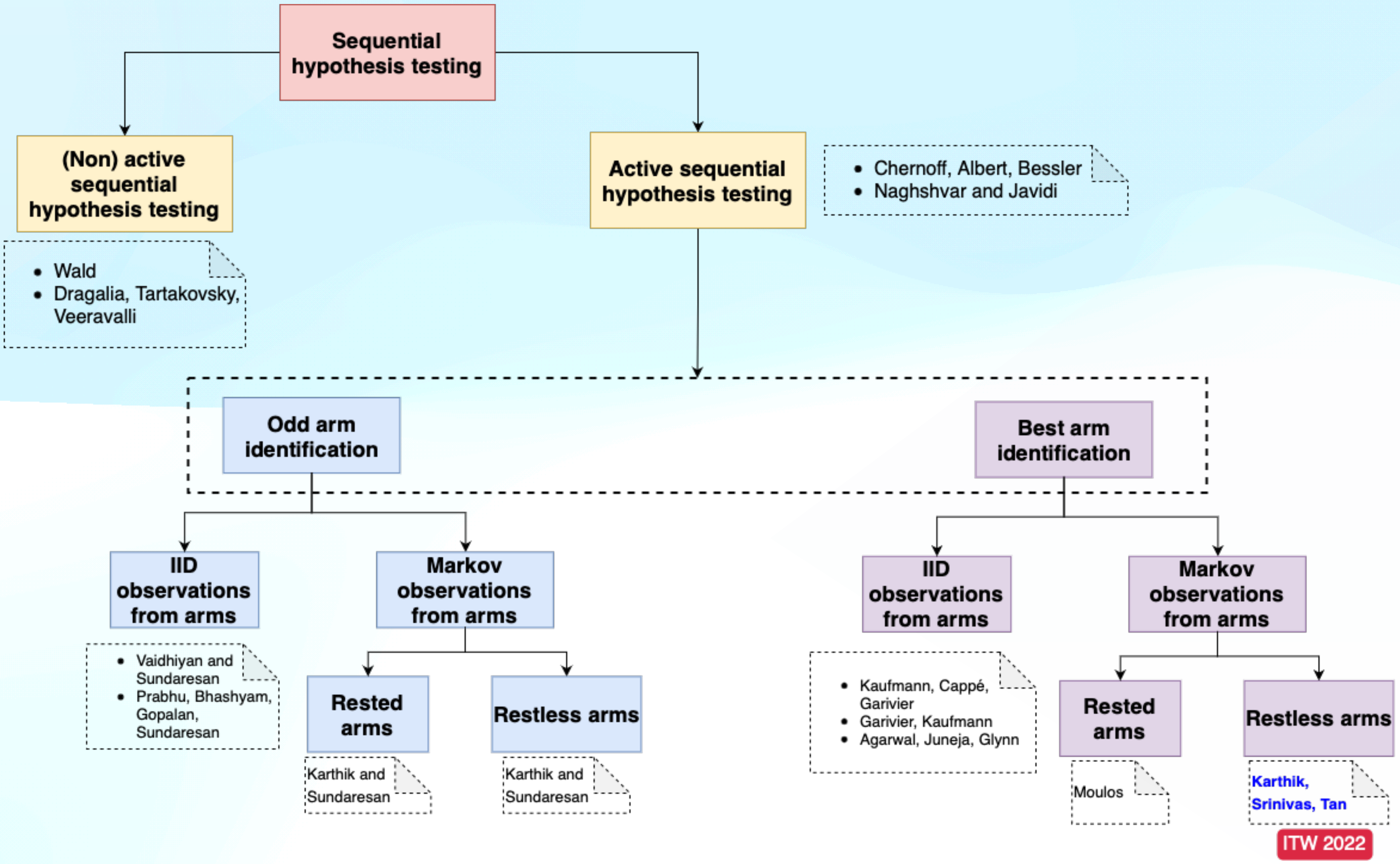
for i.i.d. arms

$$\limsup_{R \to \infty} T_R^{\star}(P_1, \ldots, P_K) = T^{\star}(P_1, \ldots, P_K)$$

$$\frac{1}{T^{\star}(P_1, \ldots, P_K)} \leq \liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}^{\pi}[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \lim_{R \to \infty} \limsup_{\delta \downarrow 0} \frac{E^{\pi^{\star}}[\text{stopping time under } \pi^{\star}]}{\log(1/\delta)} \leq \frac{1}{\lim_{R \to \infty} T_R^{\star}(P_1, \ldots, P_K)}$$

WHERE WE STAND

RELATED WORK



FUTURE WORK

FUTURE WORK

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \leq \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under }\pi]}{\log(1/\delta)} \leq \lim_{R\to\infty} \limsup_{\delta\downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under }\pi^\star]}{\log(1/\delta)} \leq \frac{1}{\lim_{R\to\infty} T^\star_R(P_1,\ldots,P_K)}$$

$$T^\star(C) = \sup_{\nu} \min_{C'\in Alt(C)} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\ k_{CC'}(\underline{d},\underline{i},a)$$

NUS National University of Singapore

FUTURE WORK

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \lim_{R\to\infty} \limsup_{\delta\downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)}$$

$$T^\star(C) = \sup_{\nu} \min_{C'\in Alt(C)} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\; k_{CC'}(\underline{d},\underline{i},a)$$

$\nu(\underline{d},\underline{i},a) \ge 0$ for all $(\underline{d},\underline{i},a)$

$$\sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a) = 1$$

FUTURE WORK

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \leq \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \lim_{R\to\infty} \limsup_{\delta\downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)}$$

$$T^\star(C) = \sup_{\nu} \min_{C'\in Alt(C)} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, k_{CC'}(\underline{d},\underline{i},a)$$

flow constraint

$$\left.\sum_{a=1}^{K} \nu(\underline{d}',\underline{i}',a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, \mathbb{P}((\underline{d},\underline{i},a)\to(\underline{d}',\underline{i}')) \quad \text{for all } (\underline{d}',\underline{i}')\in\mathbb{S}\right\}$$

flow out

flow in

$(\underline{d}',\underline{i}')$

$\nu(\underline{d},\underline{i},a) \geq 0$ for all $(\underline{d},\underline{i},a)$

$$\sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a) = 1$$

FUTURE WORK

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \leq \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \lim_{R\to\infty} \limsup_{\delta\downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)}$$

$$T^\star(C) = \sup_{\nu} \min_{C'\in Alt(C)} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, k_{CC'}(\underline{d},\underline{i},a)$$

flow constraint

$$\left.\sum_{a=1}^{K} \nu(\underline{d}',\underline{i}',a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, \mathbb{P}((\underline{d},\underline{i},a)\to(\underline{d}',\underline{i}')) \quad \text{for all } (\underline{d}',\underline{i}')\in\mathbb{S}\right\}$$

flow out

$(\underline{d}',\underline{i}')$

flow in

$$\nu(\underline{d},\underline{i},a) = \mu(\underline{d},\underline{i})\cdot\lambda(a\,|\,\underline{d},\underline{i})$$

Ergodic state-action occupancy measures

$\nu(\underline{d},\underline{i},a) \geq 0$ for all $(\underline{d},\underline{i},a)$

$$\sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a) = 1$$

NUS National University of Singapore

FUTURE WORK

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \le \liminf_{\delta\downarrow 0}\ \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \le \lim_{R\to\infty}\ \limsup_{\delta\downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \le \frac{1}{\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)}$$

$$T^\star(C) = \sup_{\nu}\ \min_{C'\in Alt(C)} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\ k_{CC'}(\underline{d},\underline{i},a)$$

flow constraint

$$\left.\sum_{a=1}^{K}\nu(\underline{d}',\underline{i}',a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}}\sum_{a=1}^{K}\nu(\underline{d},\underline{i},a)\ \mathbb{P}((\underline{d},\underline{i},a)\to(\underline{d}',\underline{i}')) \quad \text{for all } (\underline{d}',\underline{i}')\in\mathbb{S}\right\}$$

flow out

$(\underline{d}',\underline{i}')$

flow in

??

$$\nu(\underline{d},\underline{i},a) = \mu(\underline{d},\underline{i})\cdot\lambda(a\,|\,\underline{d},\underline{i})$$

Ergodic state-action occupancy measures

$\nu(\underline{d},\underline{i},a)\ge 0$ for all $(\underline{d},\underline{i},a)$

$$\sum_{(\underline{d},\underline{i})\in\mathbb{S}}\sum_{a=1}^{K}\nu(\underline{d},\underline{i},a) = 1$$

NUS National University of Singapore

FUTURE WORK

$$\frac{1}{T^\star(P_1,\ldots,P_K)} \leq \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}^\pi[\text{stopping time under } \pi]}{\log(1/\delta)} \leq \lim_{R\to\infty} \limsup_{\delta\downarrow 0} \frac{E^{\pi^\star}[\text{stopping time under } \pi^\star]}{\log(1/\delta)} \leq \frac{1}{\lim_{R\to\infty} T_R^\star(P_1,\ldots,P_K)}$$

$$T^\star(C) = \sup_{\nu} \min_{C'\in Alt(C)} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\; k_{CC'}(\underline{d},\underline{i},a)$$

flow constraint

$$\left.\sum_{a=1}^{K} \nu(\underline{d}',\underline{i}',a) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, \mathbb{P}((\underline{d},\underline{i},a)\to(\underline{d}',\underline{i}')) \quad \text{for all } (\underline{d}',\underline{i}')\in\mathbb{S} \right\}$$

flow out

$(\underline{d}',\underline{i}')$

flow in

??

$$\nu(\underline{d},\underline{i},a) = \mu(\underline{d},\underline{i})\cdot\lambda(a\,|\,\underline{d},\underline{i})$$

Ergodic state-action occupancy measures

$\nu(\underline{d},\underline{i},a) \geq 0$ for all $(\underline{d},\underline{i},a)$

$$\sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a) = 1$$

- Unknown TPMs
- Hidden Markov observations
- Second-order asymptotics

NUS National University of Singapore

ACKNOWLEDGEMENTS

- Srinivas and Vincent for their enthusiasm
- National Research Foundation, Singapore, and DSO National Laboratories for the generous sponsorship (award no: AISG2-RP-2020-018)

Srinivas Reddy Kota
ksreddy@nus.edu.sg
https://sites.google.com/view/srinivas-reddy-kota/home

Vincent Y. F. Tan
vtan@nus.edu.sg
https://vyftan.github.io/

NUS National University of Singapore

ACKNOWLEDGEMENTS

- Srinivas and Vincent for their enthusiasm
- National Research Foundation, Singapore, and DSO National Laboratories for the generous sponsorship (award no: AISG2-RP-2020-018)

Srinivas Reddy Kota
ksreddy@nus.edu.sg
https://sites.google.com/view/srinivas-reddy-kota/home

Vincent Y. F. Tan
vtan@nus.edu.sg
https://vyftan.github.io/

THANK YOU!

NUS National University of Singapore

ACKNOWLEDGEMENTS

- Srinivas and Vincent for their enthusiasm
- National Research Foundation, Singapore, and DSO National Laboratories for the generous sponsorship (award no: AISG2-RP-2020-018)

Srinivas Reddy Kota
ksreddy@nus.edu.sg
https://sites.google.com/view/srinivas-reddy-kota/home

Vincent Y. F. Tan
vtan@nus.edu.sg
https://vyftan.github.io/

THANK YOU!

Questions? Let's talk!

Contact: karthik@nus.edu.sg
https://karthikpn.com

NUS National University of Singapore