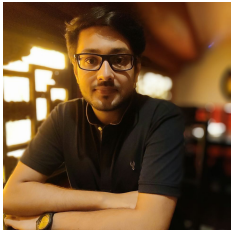
Almost-Optimal Best Restless Markov Arm Identification with Fixed Confidence

P. N. Karthik

Institute of Data Science
National University of Singapore
August 17, 2023

Joint Work With

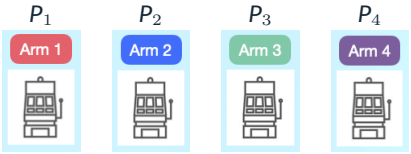


Arpan Mukherjee
RPI, New York

Vincent Tan
NUS, Singapore

Ali Tajer
RPI, New York

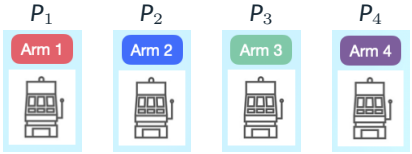
Problem Setup



Problem Setup

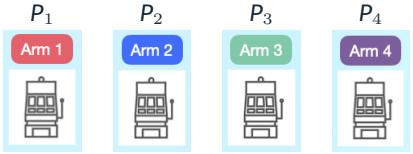
P_1	P_2	P_3	P_4
Arm 1	Arm 2	Arm 3	Arm 4
			
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$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

Problem Setup



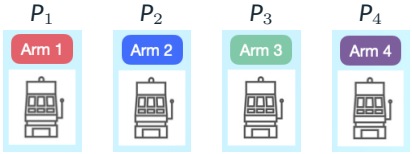
Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$

Problem Setup



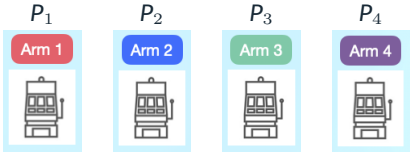
Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$

Problem Setup



Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$
2	3	$X_{2,3}$

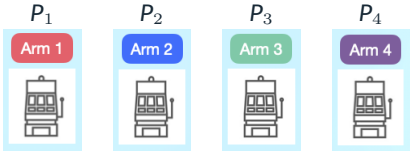
Problem Setup



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

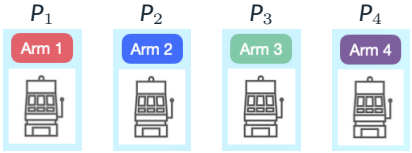
Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$
2	3	$X_{2,3}$
3	4	$X_{3,4}$

Problem Setup



Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$
2	3	$X_{2,3}$
3	4	$X_{3,4}$
4	3	$X_{4,3}$
5	3	$X_{5,4}$
6	2	$X_{6,2}$
7	1	$X_{7,1}$

Problem Setup

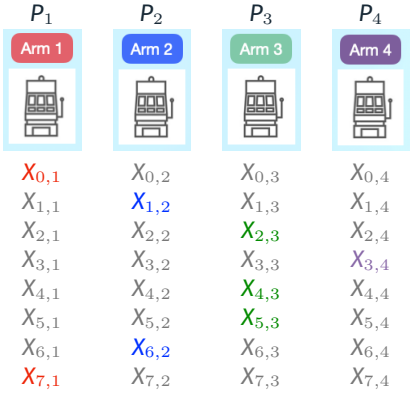


P_1	P_2	P_3	P_4
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$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

Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$
2	3	$X_{2,3}$
3	4	$X_{3,4}$
4	3	$X_{4,3}$
5	3	$X_{5,4}$
6	2	$X_{6,2}$
7	1	$X_{7,1}$

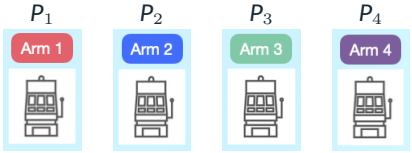
Goal: fixed-confidence best arm identification

Problem Setup



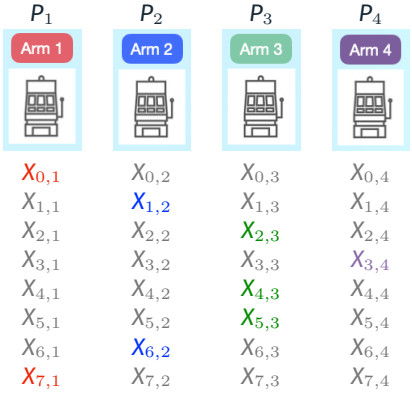
- K-armed bandit, $K \geq 2$

Problem Setup



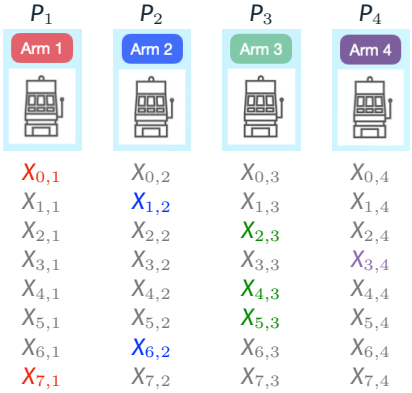
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$

Problem Setup



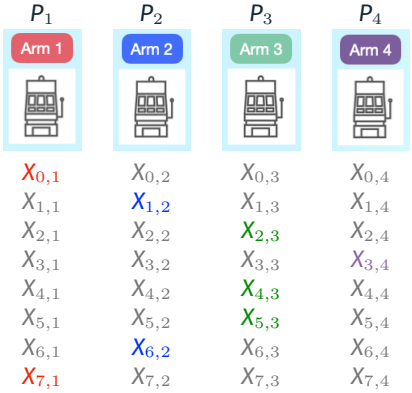
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite

Problem Setup



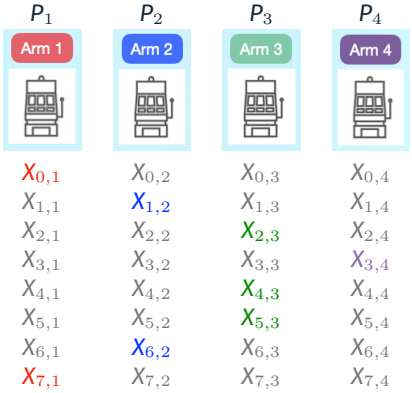
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms

Problem Setup



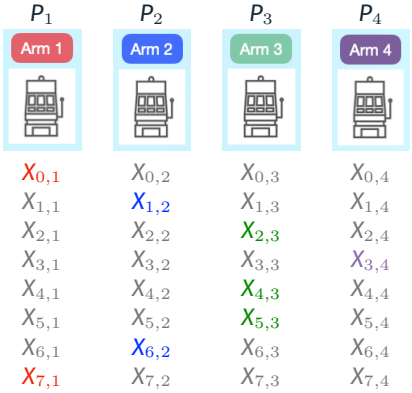
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms
- P_a : TPM of arm a, ergodic

Problem Setup



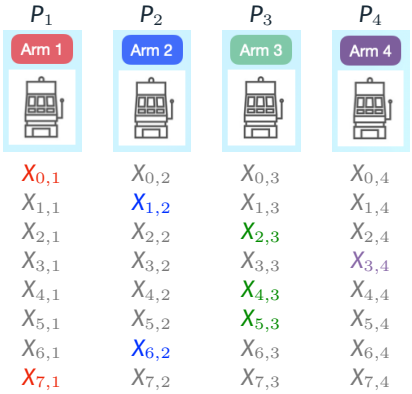
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms
- P_a : TPM of arm a, ergodic
- μ_a : stationary distribution of arm a

Problem Setup



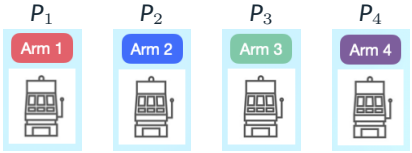
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms
- P_a : TPM of arm a, ergodic
- μ_a : stationary distribution of arm a
- $\{P_a, \mu_a : a \in [K]\}$ unknown

Problem Setup



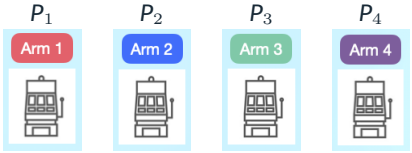
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms
- P_a : TPM of arm a, ergodic
- μ_a : stationary distribution of arm a
- $\{P_a, \mu_a : a \in [K]\}$ unknown
- Known reward function $f : \mathcal{S} \to \mathbb{R}$

Problem Setup



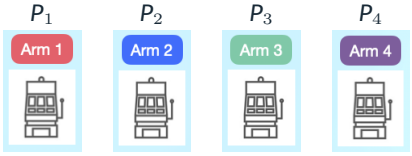
- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms
- P_a : TPM of arm a, ergodic
- μ_a : stationary distribution of arm a
- $\{P_a, \mu_a : a \in [K]\}$ unknown
- Known reward function $f : \mathcal{S} \to \mathbb{R}$
- $\eta_a = \langle f, \mu_a \rangle = \sum_{i \in \mathcal{S}} f(i)\, \mu_a(i)$

Problem Setup



- K-armed bandit, $K \geq 2$
- Arm a: homogeneous DTMC $\{X_{t,a}\}_{t \geq 0}$
- State space of each arm: $\mathcal{S}$, finite
- Restless arms
- P_a : TPM of arm a, ergodic
- μ_a : stationary distribution of arm a
- $\{P_a, \mu_a : a \in [K]\}$ unknown
- Known reward function $f : \mathcal{S} \to \mathbb{R}$
- $\eta_a = \langle f, \mu_a \rangle = \sum_{i \in \mathcal{S}} f(i)\, \mu_a(i)$
- Best arm $a^\star = \arg\max_a \eta_a$

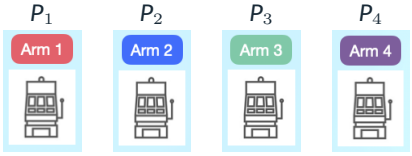
Objective



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

π - policy for best arm identification
τ_π - stopping time of policy π

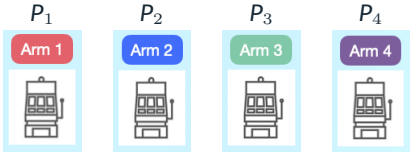
Objective



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

π - policy for best arm identification
τ_π - stopping time of policy π

Objective

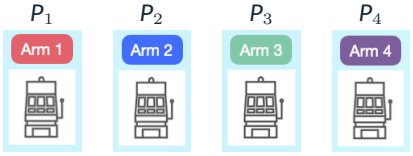


P_1	P_2	P_3	P_4
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$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

π - policy for best arm identification
τ_π - stopping time of policy π

1. $\Pr(\tau_\pi < +\infty) = 1$

Objective



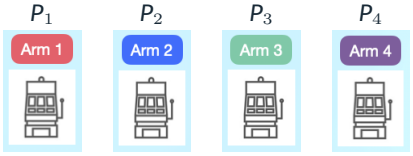
π - policy for best arm identification
τ_π - stopping time of policy π

1. $\Pr(\tau_\pi < +\infty) = 1$
2. Fixed-confidence setting: given any $\delta \in (0,1)$,

Instance	Arm 1	Arm 4	Other arms
P_1 P_2 P_3 P_4	$\geq 1-\delta$	$\leq \delta$	$\leq \delta$
Q_1 Q_2 Q_3 Q_4	$\leq \delta$	$\geq 1-\delta$	$\leq \delta$

$$\Pr(a_\pi \neq a^\star) \leq \delta \quad \forall \text{ instances}$$

Objective



π - policy for best arm identification
τ_π - stopping time of policy π

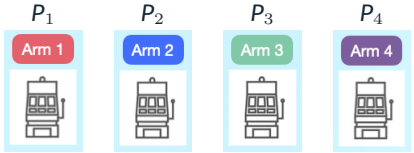
1. $\Pr(\tau_\pi < +\infty) = 1$
2. Fixed-confidence setting: given any $\delta \in (0,1)$,

Instance	Arm 1	Arm 4	Other arms
$\boxed{P_1}\ P_2\ P_3\ P_4$	$\geq 1-\delta$	$\leq \delta$	$\leq \delta$
$Q_1\ Q_2\ Q_3\ \boxed{Q_4}$	$\leq \delta$	$\geq 1-\delta$	$\leq \delta$

$\Pr(a_\pi \neq a^\star) \leq \delta \quad \forall$ instances

3. $\Pi(\delta) = \{\pi$ satisfying 1. and 2.$\}$

Objective



π - policy for best arm identification
τ_π - stopping time of policy π

1. $\Pr(\tau_\pi < +\infty) = 1$
2. Fixed-confidence setting: given any $\delta \in (0,1)$,

Instance	Arm 1	Arm 4	Other arms
P_1 P_2 P_3 P_4	$\geq 1-\delta$	$\leq \delta$	$\leq \delta$
Q_1 Q_2 Q_3 Q_4	$\leq \delta$	$\geq 1-\delta$	$\leq \delta$

$\Pr(a_\pi \neq a^\star) \leq \delta \quad \forall$ instances

3. $\Pi(\delta) = \{\pi$ satisfying 1. and 2.$\}$

Fixed-confidence BAI: $\inf_{\pi \in \Pi(\delta)} \mathbb{E}[\tau_\pi]$

Model: Exponential Family of TPMs

$\Theta \subset \mathbb{R}$ - known parameter space
P – irreducible TPM on $\mathcal{S}$ (generator)
Known reward function $f : \mathcal{S} \to \mathbb{R}$

$$\theta \mapsto \tilde{P}_\theta(i,j) = P(i,j)\, e^{\theta f(j)}, \quad i,j \in \mathcal{S}$$

Property of $\tilde{P}_\theta$: there exist vectors u_θ, v_θ such that

$$u_\theta(i) > 0 \quad \forall i, \qquad v_\theta(i) > 0 \quad \forall i, \qquad \tilde{P}_\theta\, u_\theta = \rho(\theta)\, u_\theta, \qquad v_\theta^\top\, \tilde{P}_\theta = \rho(\theta)\, v_\theta^\top$$

$$\theta \mapsto P_\theta(i,j) = \frac{v_\theta(j)}{\rho(\theta)\, v_\theta(i)} \times \tilde{P}_\theta(i,j), \quad i,j \in \mathcal{S}$$

Model: Exponential Family of TPMs

$\Theta \subset \mathbb{R}$ - known parameter space
P – irreducible TPM on $\mathcal{S}$ (generator)
Known reward function $f : \mathcal{S} \to \mathbb{R}$

$$\theta \mapsto \tilde{P}_\theta(i,j) = P(i,j)\, e^{\theta f(j)}, \quad i,j \in \mathcal{S}$$

Property of $\tilde{P}_\theta$: there exist vectors u_θ, v_θ such that

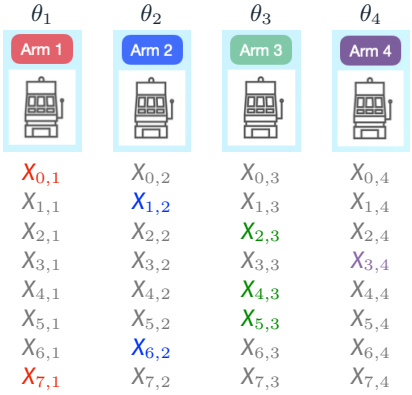
$$u_\theta(i) > 0 \quad \forall i, \qquad v_\theta(i) > 0 \quad \forall i, \qquad \tilde{P}_\theta\, u_\theta = \rho(\theta)\, u_\theta, \qquad v_\theta^\top\, \tilde{P}_\theta = \rho(\theta)\, v_\theta^\top$$

$$\theta \mapsto P_\theta(i,j) = \frac{v_\theta(j)}{\rho(\theta)\, v_\theta(i)} \times \tilde{P}_\theta(i,j), \quad i,j \in \mathcal{S}$$

Ergodic ($P_\theta \iff \mu_\theta = u_\theta \odot v_\theta$) ✔

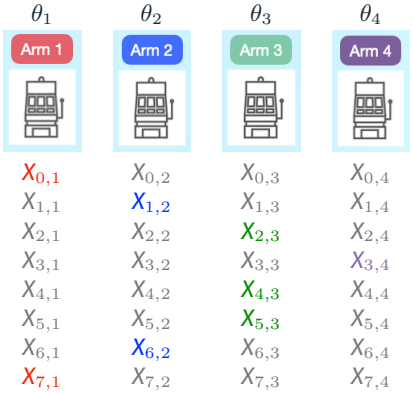
$\theta \mapsto \eta_\theta = \langle f, \mu_\theta \rangle$ strictly increasing bijection ✔

Objective



- $\boldsymbol{\theta} = [\theta_a : a \in [K]]^\top$ unknown
- $a^\star = \arg\max_a \eta_{\theta_a} = \arg\max_a \langle f, \mu_{\theta_a} \rangle$

Objective



- $\boldsymbol{\theta} = [\theta_a : a \in [K]]^\top$ unknown
- $a^\star = \arg\max_a \eta_{\theta_a} = \arg\max_a \langle f, \mu_{\theta_a} \rangle$

Fixed-confidence BAI: $\inf_{\pi \in \Pi(\delta)} \mathbb{E}[\tau_\pi]$

Markov Decision Process

Markov Decision Process

θ_1	θ_2	θ_3	θ_4
Arm 1	Arm 2	Arm 3	Arm 4
			
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$

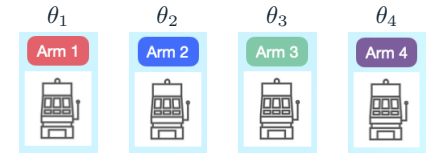
Arm delays

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

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$

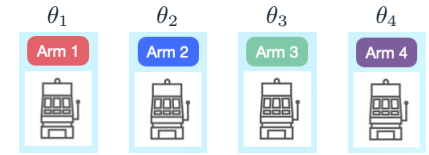
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9				

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9				

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$

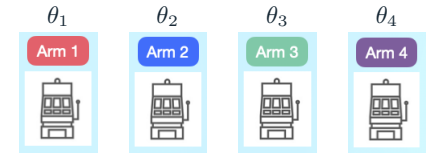
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2			

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$			

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$

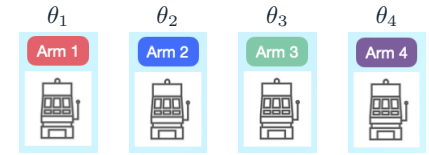
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1		

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$		

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$

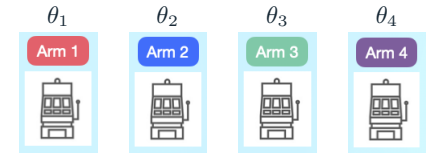
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$

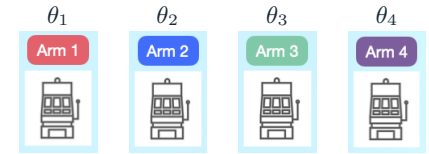
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

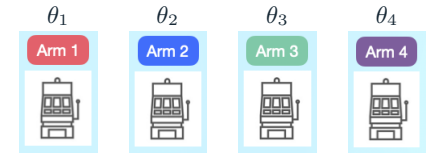
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10				

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10				

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3			

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$			

Markov Decision Process

θ_1	θ_2	θ_3	θ_4
Arm 1	Arm 2	Arm 3	Arm 4
			
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2		

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$		

Markov Decision Process

θ_1	θ_2	θ_3	θ_4
Arm 1	Arm 2	Arm 3	Arm 4
			
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

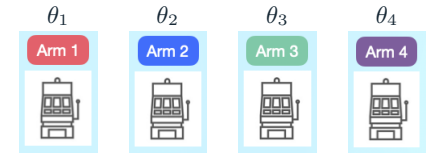
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

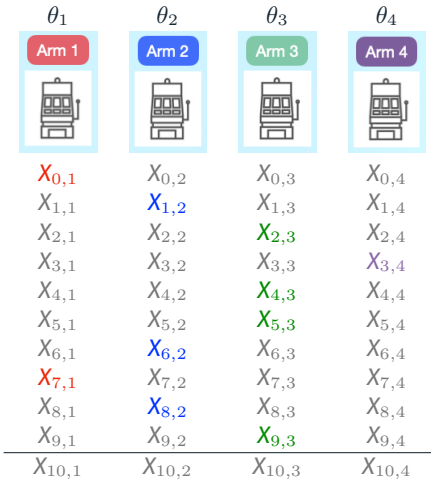
Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

Markov Decision Process



Arm delays

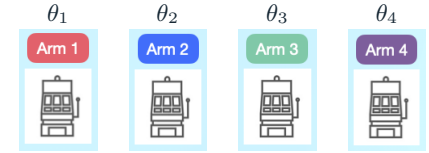
n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

$$\underline{d}(n) = [d_a(n) : a \in [K]]^\top$$
$$\underline{i}(n) = [i_a(n) : a \in [K]]^\top$$

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

$$\underline{d}(n) = [d_a(n) : a \in [K]]^\top$$
$$\underline{i}(n) = [i_a(n) : a \in [K]]^\top$$

$$(\underline{d}(n), \underline{i}(n)) \longrightarrow A_n \longrightarrow (\underline{d}(n+1), \underline{i}(n+1)) \longrightarrow A_{n+1} \longrightarrow (\underline{d}(n+2), \underline{i}(n+2)) \longrightarrow \cdots$$

Markov Decision Process

θ_1	θ_2	θ_3	θ_4
Arm 1	Arm 2	Arm 3	Arm 4
			
$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

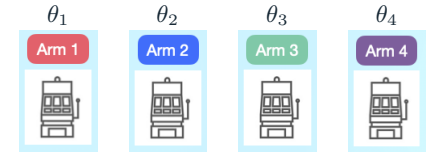
Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

$$\underline{d}(n) = [d_a(n) : a \in [K]]^\top$$
$$\underline{i}(n) = [i_a(n) : a \in [K]]^\top$$

State space: $\mathbb{S} = \{(\underline{d}, \underline{i})\} \subset \mathbb{N}^K \times \mathcal{S}^K$ countably infinite

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

Action space: $[K]$ finite

Arm delays

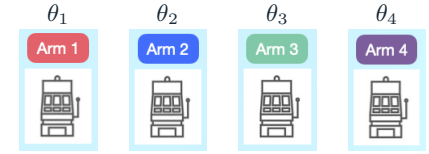
n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

$$\underline{d}(n) = [d_a(n) : a \in [K]]^\top$$
$$\underline{i}(n) = [i_a(n) : a \in [K]]^\top$$

Markov Decision Process



Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

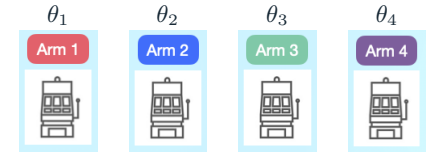
Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

$$\underline{d}(n) = [d_a(n) : a \in [K]]^\top$$
$$\underline{i}(n) = [i_a(n) : a \in [K]]^\top$$

Transition probabilities: $\Pr(\underline{d}(9), \underline{i}(9) \mid \underline{d}(8), \underline{i}(8), A_8 = 2) = \Pr(X_{8,2} | X_{6,2}) = P^2_{\theta_2}(X_{6,2}, X_{8,2})$

Markov Decision Process



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

Arm delays

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

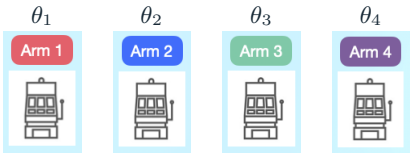
Last observed states

n	$i_1(n)$	$i_2(n)$	$i_3(n)$	$i_4(n)$
8	$X_{7,1}$	$X_{6,2}$	$X_{5,3}$	$X_{3,4}$
9	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	$X_{3,4}$
10	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	$X_{3,4}$

$$\underline{d}(n) = [d_a(n) : a \in [K]]^\top$$
$$\underline{i}(n) = [i_a(n) : a \in [K]]^\top$$

Transition probabilities: $\Pr(\underline{d}(10), \underline{i}(10) \mid \underline{d}(9), \underline{i}(9), A_9 = 3) = P^4_{\theta_3}(X_{5,3}, X_{9,3})$ TPM powers

Flow Constraint



$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$
$X_{8,1}$	$X_{8,2}$	$X_{8,3}$	$X_{8,4}$
$X_{9,1}$	$X_{9,2}$	$X_{9,3}$	$X_{9,4}$
$X_{10,1}$	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$

$$(\underline{d}(n), \underline{i}(n)) \longrightarrow A_n \xrightarrow{\text{entry}} (\underline{d}(n+1), \underline{i}(n+1)) \xrightarrow{\text{exit}} \cdots$$
$$A_{n+1} \xrightarrow{\text{entry}} (\underline{d}(n+2), \underline{i}(n+2)) \xrightarrow{\text{exit}} A_{n+2} \xrightarrow{\text{entry}} \cdots$$

$$\xrightarrow{\text{entry}} \boxed{(\underline{d}, \underline{i})} \xrightarrow{\text{exit}}$$

$$N(n, \underline{d}, \underline{i}, a) = \sum_{t=K}^{n} \mathbf{1}_{\{\underline{d}(t)=\underline{d},\ \underline{i}(t)=\underline{i},\ A_t=a\}}$$

$$N(n, \underline{d}, \underline{i}) = \sum_{a=1}^{K} N(n, \underline{d}, \underline{i}, a)$$

Flow Constraint

Under every policy π,

$$\left| \mathbb{E}[N(n, \underline{d}', \underline{i}')] - \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \mathbb{E}[N(n, \underline{d}, \underline{i}, a)] \, Q_{\theta}(\underline{d}', \underline{i}' | \underline{d}, \underline{i}, a) \right| \leq 1 \quad \forall (\underline{d}', \underline{i}') \in \mathbb{S}, \quad \forall n \geq K.$$

Maximum Delay Constraint

- Fix $R \in \mathbb{N}$ such that $R \gg K$
- Forcefully pull an arm if its delay equals R at any given time
- $\mathbb{S}_R = \{(\underline{d}, \underline{i}) \in \mathbb{S} : \max_a d_a \leq R\}, \qquad \mathbb{S}_{R,a} = \{(\underline{d}, \underline{i}) \in \mathbb{S}_R : d_a = R\}$ finite state space
- $\xrightarrow{\text{entry}}$ $(\underline{d}(n), \underline{i}(n)) \in \mathbb{S}_{R,a}$ $\xrightarrow{\text{exit}}$ $A_n = a$
- $Q_\theta \longrightarrow Q_{\theta,R}$

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

Maximum Delay Constraint

- Fix $R \in \mathbb{N}$ such that $R \gg K$
- Forcefully pull an arm if its delay equals R at any given time
- $\mathbb{S}_R = \{(\underline{d}, \underline{i}) \in \mathbb{S} : \max_a d_a \leq R\}$, $\mathbb{S}_{R,a} = \{(\underline{d}, \underline{i}) \in \mathbb{S}_R : d_a = R\}$ finite state space
- $\xrightarrow{\text{entry}}$ $(\underline{d}(n), \underline{i}(n)) \in \mathbb{S}_{R,a}$ $\xrightarrow{\text{exit}} A_n = a$
- $Q_\theta \longrightarrow Q_{\theta,R}$

n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

$R = 6$

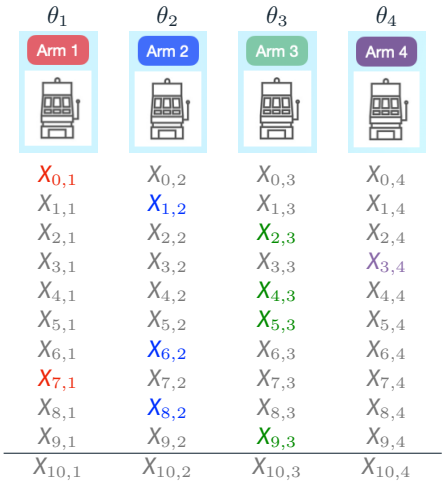
$(\underline{d}(9), \underline{i}(9)) \longrightarrow A_{10} = 4$ ✔

Flow Constraint + R-Max-Delay Constraint

$$\left| \mathbb{E}[N(n, \underline{d}', \underline{i}')] - \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \mathbb{E}[N(n, \underline{d}, \underline{i}, a)]\, Q_{\theta,R}(\underline{d}', \underline{i}' | \underline{d}, \underline{i}, a) \right| \leq 1 \quad \forall (\underline{d}', \underline{i}') \in \mathbb{S}_R, \quad \forall n \geq K,$$

$$N(n, \underline{d}, \underline{i}, a) = N(n, \underline{d}, \underline{i}) \qquad \forall (\underline{d}, \underline{i}) \in \mathbb{S}_{R,a},\ a \in [K]$$

Objective Restated



- $\boldsymbol{\theta} = [\theta_a : a \in [K]]^\top$ unknown
- $a^\star(\boldsymbol{\theta}) = \arg\max_a \eta_{\theta_a} = \arg\max_a \langle f, \mu_{\theta_a} \rangle$

Objective: $\inf_\pi \mathbb{E}[\tau_\pi]$

- ✔ $\pi \in \Pi(\delta)$
- ✔ π satisfies *R*-max-delay constraint

Converse

Converse

- Fix $\boldsymbol{\theta}$. Assume $a^\star(\boldsymbol{\theta}) = 1$
- Set of alternative instances:

$$\begin{aligned}\textsc{Alt}(\boldsymbol{\theta}) &= \{\boldsymbol{\lambda} \in \Theta^K : \exists a \neq a^\star(\boldsymbol{\theta}) \text{ such that } \eta_{\lambda_a} > \eta_{\lambda_1}\} \\ &= \{\boldsymbol{\lambda} \in \Theta^K : \exists a \neq a^\star(\boldsymbol{\theta}) \text{ such that } \lambda_a > \lambda_1\} = \bigcup_{a=1}^{K} \{\boldsymbol{\lambda} \in \Theta^K : \lambda_a > \lambda_1\}\end{aligned}$$

Converse

- Fix $\boldsymbol{\theta}$. Assume $a^\star(\boldsymbol{\theta}) = 1$
- Set of alternative instances:

$$\begin{aligned}\textsc{Alt}(\boldsymbol{\theta}) &= \{\boldsymbol{\lambda} \in \Theta^K : \exists a \neq a^\star(\boldsymbol{\theta}) \text{ such that } \eta_{\lambda_a} > \eta_{\lambda_1}\} \\ &= \{\boldsymbol{\lambda} \in \Theta^K : \exists a \neq a^\star(\boldsymbol{\theta}) \text{ such that } \lambda_a > \lambda_1\} = \bigcup_{a=1}^{K} \{\boldsymbol{\lambda} \in \Theta^K : \lambda_a > \lambda_1\}\end{aligned}$$

Proposition

$$\inf_{\pi \in \Pi(\delta)} \mathbb{E}[\tau_\pi] \geq \frac{\delta \log \frac{\delta}{1-\delta} + (1-\delta) \log \frac{1-\delta}{\delta}}{T^\star_R(\boldsymbol{\theta})},$$

$$T^\star_R(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \textsc{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, D_{KL}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d}, \underline{i}, a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d}, \underline{i}, a)),$$

$$\Sigma_R(\boldsymbol{\theta}) = \Big\{\nu : \checkmark \textit{prob. dist.}, \ \checkmark \boxed{\textit{flow constraint}}, \ \checkmark \boxed{\textit{R-max-delay constraint}}\Big\}$$

Converse – 1

$$\inf_{\pi \in \Pi(\delta)} \mathbb{E}[\tau_\pi] \geq \frac{\delta \log \frac{\delta}{1-\delta} + (1-\delta) \log \frac{1-\delta}{\delta}}{T_R^\star(\boldsymbol{\theta})}$$

Converse – 1

$$\inf_{\pi \in \Pi(\delta)} \mathbb{E}[\tau_\pi] \geq \frac{\delta \log \frac{\delta}{1-\delta} + (1-\delta) \log \frac{1-\delta}{\delta}}{T_R^\star(\boldsymbol{\theta})} \sim \frac{\log \frac{1}{\delta}}{T_R^\star(\boldsymbol{\theta})}$$

Converse – 1

$$\inf_{\pi \in \Pi(\delta)} \mathbb{E}[\tau_\pi] \geq \frac{\delta \log \frac{\delta}{1-\delta} + (1-\delta) \log \frac{1-\delta}{\delta}}{T_R^\star(\boldsymbol{\theta})} \sim \frac{\log \frac{1}{\delta}}{T_R^\star(\boldsymbol{\theta})}$$

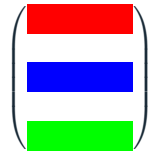
$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}[\tau_\pi]}{\log(1/\delta)} \geq \frac{1}{T_R^\star(\boldsymbol{\theta})}$$

Converse – 2

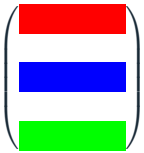
$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, D_{\mathrm{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d}, \underline{i}, a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d}, \underline{i}, a))$$

Converse – 2

$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathsf{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a))$$



$Q_{\boldsymbol{\theta},R}, \quad \boldsymbol{\theta} = [\theta_a : a \in [K]]^\top$



$Q_{\boldsymbol{\lambda},R}, \quad \boldsymbol{\lambda} = [\lambda_a : a \in [K]]^\top$

$$Q_{\boldsymbol{\theta},R}(\underline{d}', \underline{i}' \mid \underline{d}, \underline{i}, a) = \begin{cases} P_{\theta_a}^{d_a}(i_a, i_a'), & (\underline{d},\underline{i}) \xrightarrow{a} (\underline{d}', \underline{i}'), \\ 0, & \text{otherwise} \end{cases}$$

Converse – 3

$$\Sigma_R(\boldsymbol{\theta}) = \Big\{\nu : \checkmark\text{prob. dist.}, \quad \checkmark\boxed{\text{flow constraint}}, \quad \checkmark\boxed{\text{R-max-delay constraint}}\Big\}$$

$$\Sigma_R(\boldsymbol{\theta}) = \Bigg\{\nu : \quad \nu(\underline{d}, \underline{i}, a) \geq 0 \quad \forall (\underline{d}, \underline{i}, a),$$

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

$$\sum_{a=1}^{K} \nu(\underline{d}', \underline{i}', a) = \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, Q_{\boldsymbol{\theta},R}(\underline{d}', \underline{i}' | \underline{d}, \underline{i}, a) \quad \forall (\underline{d}', \underline{i}') \in \mathbb{S}_R,$$

$$\nu(\underline{d}, \underline{i}, a) = \sum_{a'=1}^{K} \nu(\underline{d}, \underline{i}, a') \quad \forall (\underline{d}, \underline{i}) \in \mathbb{S}_{R,a},\ a \in [K] \Bigg\}$$

Converse – 4

$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\text{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a))$$

$$\psi(\nu,\boldsymbol{\theta}) = \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\text{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a)), \quad \nu \in \Sigma_R(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta^K$$

Lemma

Converse – 4

$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathrm{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a))$$

$$\psi(\nu,\boldsymbol{\theta}) = \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathrm{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a)), \quad \nu \in \Sigma_R(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta^K$$

Lemma

- ψ *is continuous*

Continuity of ψ – Berge's maximum theorem for non-compact sets (Feinberg et al., 2014)

Converse – 4

$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathrm{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a))$$

$$\psi(\nu,\boldsymbol{\theta}) = \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathrm{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a)), \quad \nu \in \Sigma_R(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta^K$$

Lemma

- ψ *is continuous*
- $\boldsymbol{\theta} \mapsto \mathcal{W}^\star(\boldsymbol{\theta}) = \{\nu \in \Sigma_R(\boldsymbol{\theta}) : \psi(\nu,\boldsymbol{\theta}) = T_R^\star(\boldsymbol{\theta})\}$ *is upper-semicontinuous*

Continuity of ψ – Berge's maximum theorem for non-compact sets (Feinberg et al., 2014)

Converse – 4

$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\text{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a))$$

$$\psi(\nu,\boldsymbol{\theta}) = \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\text{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a)), \quad \nu \in \Sigma_R(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta^K$$

Lemma

- ψ *is continuous*
- $\boldsymbol{\theta} \mapsto \mathcal{W}^\star(\boldsymbol{\theta}) = \{\nu \in \Sigma_R(\boldsymbol{\theta}) : \psi(\nu,\boldsymbol{\theta}) = T_R^\star(\boldsymbol{\theta})\}$ *is upper-semicontinuous*

Continuity of ψ – Berge's maximum theorem for non-compact sets (Feinberg et al., 2014)

Key idea for achievability: $\left[\frac{N(n,\underline{d},\underline{i},a)}{n} : (\underline{d},\underline{i},a) \in \mathbb{S}_R \times [K]\right]^\top \longrightarrow \mathcal{W}^\star(\boldsymbol{\theta})$

Lower Bounds from Prior Works

Restless arms

$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, D_{\mathsf{KL}}(Q_{\boldsymbol{\theta},R}(\cdot|\underline{d}, \underline{i}, a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d}, \underline{i}, a))$$

Rested arms Moulos (2019)

$$T^\star(\boldsymbol{\theta}) = \sup_{\nu} \inf_{\boldsymbol{\lambda} \in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{i \in \mathcal{S}} \sum_{a=1}^{K} \nu(a)\, \mu_{\theta_a}(i)\, D_{\mathsf{KL}}(P_{\theta_a}(i, \cdot) \| P_{\lambda_a}(i, \cdot))$$

Independent observations from arms Garivier and Kaufmann (2016)

$$T^\star(\boldsymbol{\theta}) = \sup_{\nu} \inf_{\boldsymbol{\lambda} \in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{a=1}^{K} \nu(a)\, D_{\mathsf{KL}}(\theta_a, \lambda_a)$$

Achievability

Uniform Policy and Ergodicity

$$\pi^{\mathsf{unif}}(a|\underline{d},\underline{i}) = \begin{cases} \frac{1}{K}, & (\underline{d},\underline{i}) \in \bigcup_{a'=1}^{K} \mathbb{S}_{R,a'}, \\ 1, & (\underline{d},\underline{i}) \in \mathbb{S}_{R,a}, \\ 0, & (\underline{d},\underline{i}) \in \bigcup_{a' \neq a} \mathbb{S}_{R,a'}. \end{cases}$$

Uniform Policy and Ergodicity

$$\pi^{\text{unif}}(a|\underline{d},\underline{i}) = \begin{cases} \frac{1}{K}, & (\underline{d},\underline{i}) \in \bigcup_{a'=1}^{K} \mathbb{S}_{R,a'}, \\ 1, & (\underline{d},\underline{i}) \in \mathbb{S}_{R,a}, \\ 0, & (\underline{d},\underline{i}) \in \bigcup_{a' \neq a} \mathbb{S}_{R,a'}. \end{cases}$$

Transition kernel under $\boldsymbol{\theta} \in \Theta^K$ and π^{unif}:

$$Q_{\boldsymbol{\theta},\pi^{\text{unif}}}(\underline{d}',\underline{i}'|\underline{d},\underline{i}) = \sum_{a=1}^{K} Q_{\boldsymbol{\theta},R}(\underline{d}',\underline{i}'|\underline{d},\underline{i},a) \cdot \pi^{\text{unif}}(a|\underline{d},\underline{i}) \quad \forall (\underline{d},\underline{i}), (\underline{d}',\underline{i}') \in \mathbb{S}_R$$

Uniform Policy and Ergodicity

$$\pi^{\text{unif}}(a|\underline{d},\underline{i}) = \begin{cases} \frac{1}{K}, & (\underline{d},\underline{i}) \in \bigcup_{a'=1}^{K} \mathbb{S}_{R,a'}, \\ 1, & (\underline{d},\underline{i}) \in \mathbb{S}_{R,a}, \\ 0, & (\underline{d},\underline{i}) \in \bigcup_{a' \neq a} \mathbb{S}_{R,a'}. \end{cases}$$

Transition kernel under $\boldsymbol{\theta} \in \Theta^K$ and π^{unif}:

$$Q_{\boldsymbol{\theta},\pi^{\text{unif}}}(\underline{d}',\underline{i}'|\underline{d},\underline{i}) = \sum_{a=1}^{K} Q_{\boldsymbol{\theta},R}(\underline{d}',\underline{i}'|\underline{d},\underline{i},a) \cdot \pi^{\text{unif}}(a|\underline{d},\underline{i}) \quad \forall (\underline{d},\underline{i}), (\underline{d}',\underline{i}') \in \mathbb{S}_R$$

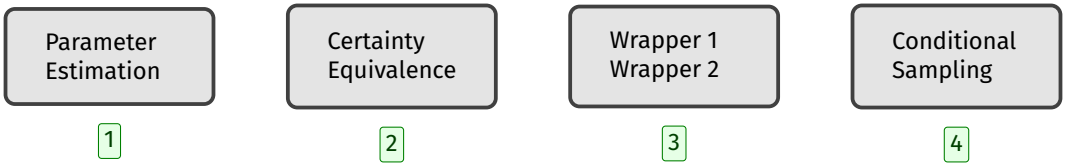
Lemma

$Q_{\boldsymbol{\theta},\pi^{unif}}$ is ergodic for all $\boldsymbol{\theta} \in \Theta^K$

Stationary distribution of $Q_{\boldsymbol{\theta},\pi^{\text{unif}}} = \mu_{\boldsymbol{\theta}}^{\text{unif}} > 0$,

$$\nu_{\boldsymbol{\theta}}^{\text{unif}}(\underline{d},\underline{i},a) = \mu_{\boldsymbol{\theta}}^{\text{unif}}(\underline{d},\underline{i}) \cdot \pi^{\text{unif}}(a|\underline{d},\underline{i}) \qquad \forall (\underline{d},\underline{i},a) \in \mathbb{S}_R \times [K]$$

Achievability – Arm Selection Rule



Arm Selection Rule – 1

Parameter Estimation

Certainty Equivalence

Wrapper 1
Wrapper 2

Conditional Sampling

1

2

3

4

$\hat{\boldsymbol{\theta}}(n)$

$$N_a(n) = \sum_{t=0}^{n} \mathbf{1}_{\{A_t = a\}}$$

$$\hat{\eta}^a(n) = \begin{cases} 0, & N_a(n) = 0, \\ \frac{1}{N_a(n)} \sum_{t=0}^{n} \mathbf{1}_{\{A_t = a\}} f(\bar{X}_t), & N_a(n) > 0 \end{cases}$$

$$\hat{\eta}_a(n) \Longleftrightarrow \hat{\theta}_a(n)$$

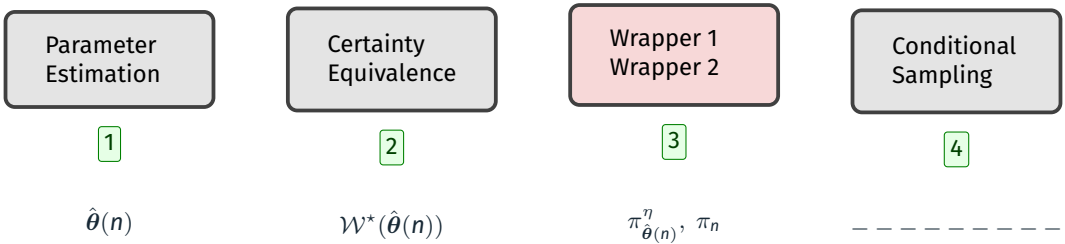
$$\hat{\boldsymbol{\theta}}(n) = [\hat{\theta}_a(n) : a \in [K]]$$

Arm Selection Rule – 2

Parameter Estimation	Certainty Equivalence	Wrapper 1 Wrapper 2	Conditional Sampling
1	2	3	4
$\hat{\boldsymbol{\theta}}(n)$	$\mathcal{W}^\star(\hat{\boldsymbol{\theta}}(n))$	– – – – – – – – –	– – – – – – – – –

$$\mathcal{W}^\star(\hat{\boldsymbol{\theta}}(n)) = \underset{\nu \in \Sigma_R(\hat{\boldsymbol{\theta}}(n))}{\arg\sup} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\hat{\boldsymbol{\theta}}(n))} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\text{KL}}(Q_{\hat{\boldsymbol{\theta}}(n),R}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d},\underline{i},a))$$

Arm Selection Rule – 3



Fix $\eta \in (0,1)$ and $\varepsilon_n \to 0$.
Pick $\nu_n^\star \in \mathcal{W}^\star(\hat{\theta}(n))$

Wrapper 1

$$\pi^\eta_{\hat{\theta}(n)}(a|\underline{d},\underline{i}) = \frac{\eta\, \nu^{\text{unif}}_{\hat{\theta}(n)}(\underline{d},\underline{i},a) + (1-\eta)\, \nu_n^\star(\underline{d},\underline{i},a)}{\eta\, \mu^{\text{unif}}_{\hat{\theta}(n)}(\underline{d},\underline{i}) + (1-\eta)\, \sum_{a'=1}^{K} \nu_n^\star(\underline{d},\underline{i},a')}, \quad (\underline{d},\underline{i},a) \in \mathbb{S}_R \times [K]$$

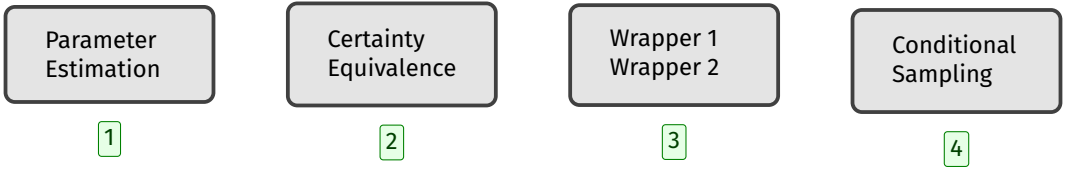
Wrapper 2

$$\pi_n = \varepsilon_n \pi^{\text{unif}} + (1-\varepsilon_n)\, \pi^\eta_{\hat{\theta}(n-1)} \quad \forall n \geq K$$

Arm Selection Rule – 4

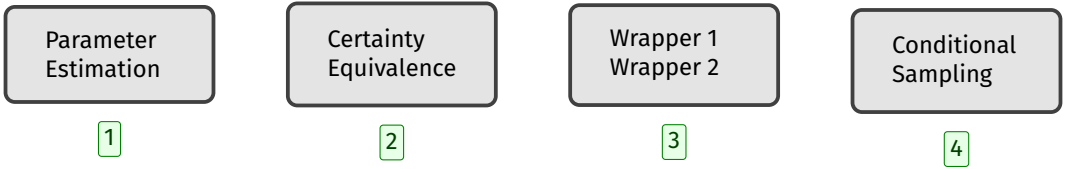
Parameter Estimation	Certainty Equivalence	Wrapper 1 Wrapper 2	Conditional Sampling
1	2	3	4
$\hat{\theta}(n)$	$\mathcal{W}^\star(\hat{\theta}(n))$	$\pi^{\eta}_{\hat{\theta}(n)},\ \pi_n$	$A_n \sim \pi_n(\cdot \vert \underline{d}(n), \underline{i}(n))$

Achievability – Arm Selection Rule



Lemma

Achievability – Arm Selection Rule



Lemma

For $S_R = |\mathbb{S}_R|$ *and* $\varepsilon_n = n^{-\frac{1}{2(1+S_R)}}$,

$$\lim_{n\to\infty} d_\infty \left(\left[\frac{N(n, \underline{d}, \underline{i}, a)}{n} : (\underline{d}, \underline{i}, a) \in \mathbb{S}_R \times [K] \right]^\top, \mathcal{W}_\eta^\star(\boldsymbol{\theta}) \right) = 0 \quad a.s.,$$

where $\mathcal{W}_\eta^\star(\boldsymbol{\theta}) = \{\eta\, \nu_{\boldsymbol{\theta}}^{unif} + (1-\eta)\, \nu : \nu \in \mathcal{W}^\star(\boldsymbol{\theta})\}$

Achievability – Stopping Rule

Empirical transition matrix:

$$\widehat{Q}_n(\underline{d}', \underline{i}' | \underline{d}, \underline{i}, a) = \begin{cases} \dfrac{1}{N(n, \underline{d}, \underline{i}, a)} \sum_{t=K}^{n} \mathbf{1}_{\{(\underline{d}(t), \underline{i}(t)) = (\underline{d}, \underline{i}),\, A_t = a,\, (\underline{d}(t+1), \underline{i}(t+1)) = (\underline{d}', \underline{i}')\}}, & N(n, \underline{d}, \underline{i}, a) > 0, \\ \dfrac{1}{S_R}, & N(n, \underline{d}, \underline{i}, a) = 0 \end{cases}$$

Achievability – Stopping Rule

Empirical transition matrix:

$$\widehat{Q}_n(\underline{d}', \underline{i}' | \underline{d}, \underline{i}, a) = \begin{cases} \dfrac{1}{N(n, \underline{d}, \underline{i}, a)} \sum_{t=K}^{n} \mathbf{1}_{\{(\underline{d}(t), \underline{i}(t)) = (\underline{d}, \underline{i}),\, A_t = a,\, (\underline{d}(t+1), \underline{i}(t+1)) = (\underline{d}', \underline{i}')\}}, & N(n, \underline{d}, \underline{i}, a) > 0, \\ \dfrac{1}{S_R}, & N(n, \underline{d}, \underline{i}, a) = 0 \end{cases}$$

Test statistic

$$Z(n) = \inf_{\lambda \in \mathrm{Alt}(\hat{\theta}(n))} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} N(n, \underline{d}, \underline{i}, a)\, D_{\mathrm{KL}}(\hat{Q}_n(\cdot | \underline{d}, \underline{i}, a) \| Q_{\lambda, R}(\cdot | \underline{d}, \underline{i}, a))$$

Achievability – Stopping Rule

$$Z(n) = \inf_{\boldsymbol{\lambda} \in \text{Alt}(\hat{\boldsymbol{\theta}}(n))} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} N(n, \underline{d}, \underline{i}, a)\, D_{\text{KL}}(\hat{Q}_n(\cdot|\underline{d}, \underline{i}, a) \| Q_{\boldsymbol{\lambda},R}(\cdot|\underline{d}, \underline{i}, a))$$

Threshold:

$$\zeta(n, \delta) = \log\left(\frac{1}{\delta}\right) + (S_R - 1) \sum_{(\underline{d},\underline{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \log\left(e\left[1 + \frac{N(n, \underline{d}, \underline{i}, a)}{S_R - 1}\right]\right)$$

Stopping rule:

$$\tau_\delta = \inf\{n \geq K : Z(n) \geq \zeta(n, \delta)\}$$

Recommendation rule:

$$\hat{a} = \arg\max_a \hat{\eta}_a(n)$$

Performance

- Error prob. $\leq \delta$
- Almost sure upper bound

$$\limsup_{\delta \downarrow 0} \frac{\tau_\delta}{\log(1/\delta)} \leq \frac{1}{(1-\eta)\, T_R^\star(\boldsymbol{\theta})} \quad \text{a.s.}$$

- Upper bound in expectation

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{(1-\eta)\, T_R^\star(\boldsymbol{\theta})}$$

$$\limsup_{\eta \downarrow 0} \limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{T_R^\star(\boldsymbol{\theta})}$$

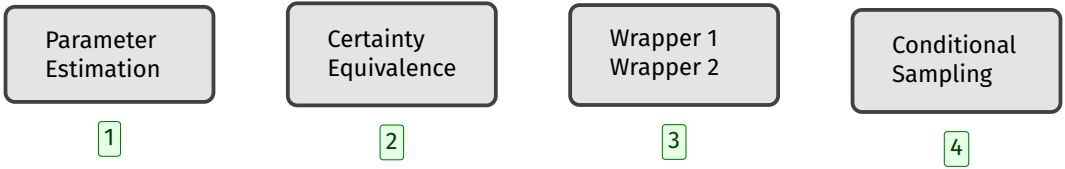
Algorithm 1 Restless Tracking

Input: $K, R, \eta, \delta > 0$
Output: $\hat{a} \in [K]$ (best arm).

Initialization: $n = 0$, $N_a(n) = 0$, $\hat{\eta}_a(n) = 0$ for all $a \in [K]$, $N(n, \underline{d}, \underline{i}, a) = 0$ for all $(\underline{d}, \underline{i}, a) \in \mathbb{S}_R \times [K]$, $\texttt{stop} = 0$

for $n < K$ **do**
  Select arm $A_n = n + 1$.
end for
while $\texttt{stop} == 0$ **do**
  Update $(\underline{d}(n), \underline{i}(n))$; update $\hat{\eta}_a(n)$ for each $a \in [K]$
  $\hat{\boldsymbol{\eta}}(n) \iff \hat{\boldsymbol{\theta}}(n)$
  Evaluate $Z(n)$
  if $Z(n) \geq \zeta(n, \delta)$ **then**
    $\texttt{stop} = 1$
    $\tau_\delta = n$
    $\hat{a} = \arg\max_a \hat{\eta}_a(n)$
  else
    Select $A_n \sim \pi_n(\cdot | \underline{d}(n), \underline{i}(n))$
    $n \leftarrow n + 1$.
  end if
end while
return $\hat{a}$.

Main Result



Theorem

$$\frac{1}{T_R^\star(\boldsymbol{\theta})} \leq \liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}[\tau_\pi]}{\log(1/\delta)} \leq \limsup_{\eta \downarrow 0} \limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{T_R^\star(\boldsymbol{\theta})}$$

More Musings

More Musings

Lemma (Montonicity)

$T^\star_R(\boldsymbol{\theta}) \leq T^\star_{R+1}(\boldsymbol{\theta})$ *for all* R

More Musings

Lemma (Montonicity)

$T_R^\star(\boldsymbol{\theta}) \leq T_{R+1}^\star(\boldsymbol{\theta})$ *for all* R

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta})$ exists, $\quad \lim_{R\to\infty} \Sigma_R(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) = \left\{\nu : \checkmark \text{prob. dist.}, \quad \checkmark \boxed{\text{flow constraint}}\right\}$

More Musings

Lemma (Montonicity)

$T_R^\star(\boldsymbol{\theta}) \leq T_{R+1}^\star(\boldsymbol{\theta})$ for all R

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta})$ exists, $\quad \lim_{R\to\infty} \Sigma_R(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) = \left\{\nu : \checkmark\text{prob. dist.}, \quad \checkmark\boxed{\text{flow constraint}}\right\}$

$$T^\star(\boldsymbol{\theta}) = \sup_{\nu\in\Sigma(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda}\in \text{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\text{KL}}(Q_{\boldsymbol{\theta}}(\cdot|\underline{d},\underline{i},a)\|Q_{\boldsymbol{\lambda}}(\cdot|\underline{d},\underline{i},a))$$

More Musings

Lemma (Montonicity)

$T_R^\star(\boldsymbol{\theta}) \leq T_{R+1}^\star(\boldsymbol{\theta})$ *for all* R

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta})$ exists, $\quad \lim_{R\to\infty} \Sigma_R(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) = \left\{\nu : \checkmark \text{prob. dist.}, \quad \checkmark \boxed{\text{flow constraint}}\right\}$

$$T^\star(\boldsymbol{\theta}) = \sup_{\nu\in\Sigma(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda}\in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathsf{KL}}(Q_{\boldsymbol{\theta}}(\cdot|\underline{d},\underline{i},a)\|Q_{\boldsymbol{\lambda}}(\cdot|\underline{d},\underline{i},a))$$

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta}) = T^\star(\boldsymbol{\theta})$??

More Musings

Lemma (Montonicity)

$T_R^\star(\boldsymbol{\theta}) \leq T_{R+1}^\star(\boldsymbol{\theta})$ for all R

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta})$ exists, $\quad \lim_{R\to\infty} \Sigma_R(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) = \Big\{\nu : \checkmark$prob. dist., $\checkmark$ flow constraint $\Big\}$

$$T^\star(\boldsymbol{\theta}) = \sup_{\nu\in\Sigma(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda}\in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathsf{KL}}(Q_{\boldsymbol{\theta}}(\cdot|\underline{d},\underline{i},a)\|Q_{\boldsymbol{\lambda}}(\cdot|\underline{d},\underline{i},a))$$

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta}) = T^\star(\boldsymbol{\theta})$??

Independent observations from arms ✔

Restless arms – OPEN

More Musings

Lemma (Montonicity)

$T_R^\star(\boldsymbol{\theta}) \le T_{R+1}^\star(\boldsymbol{\theta})$ for all R

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta})$ exists, $\quad \lim_{R\to\infty} \Sigma_R(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) = \left\{\nu : \checkmark\text{prob. dist.}, \quad \checkmark\boxed{\text{flow constraint}}\right\}$

$$T^\star(\boldsymbol{\theta}) = \sup_{\nu\in\Sigma(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda}\in \mathsf{Alt}(\boldsymbol{\theta})} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, D_{\mathsf{KL}}(Q_{\boldsymbol{\theta}}(\cdot|\underline{d},\underline{i},a) \| Q_{\boldsymbol{\lambda}}(\cdot|\underline{d},\underline{i},a))$$

$\lim_{R\to\infty} T_R^\star(\boldsymbol{\theta}) = T^\star(\boldsymbol{\theta})$??

Independent observations from arms ✔

Restless arms – OPEN

$$\frac{1}{T^\star(\boldsymbol{\theta})} \le \frac{1}{T_R^\star(\boldsymbol{\theta})} \le \liminf_{\delta\downarrow 0} \inf_{\pi\in\Pi(\delta)} \frac{\mathbb{E}[\tau_\pi]}{\log(1/\delta)} \le \limsup_{\eta\downarrow 0} \limsup_{\delta\downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \le \frac{1}{T_R^\star(\boldsymbol{\theta})} \overset{??}{\le} \frac{1}{T^\star(\boldsymbol{\theta})}$$

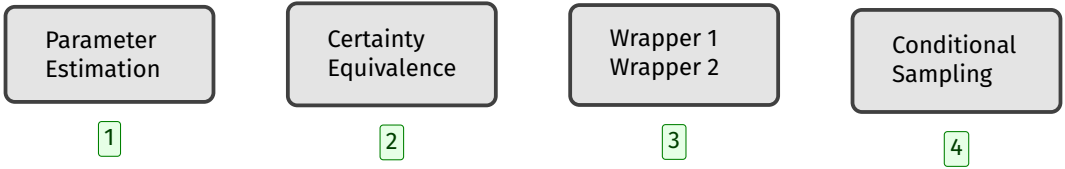
References

Feinberg, E. A., Kasyanov, P. O., and Voorneveld, M. (2014). Berge's maximum theorem for noncompact image sets. *Journal of Mathematical Analysis and Applications*, 413(2):1040–1046.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR.

Moulos, V. (2019). Optimal best Markovian arm identification with fixed confidence. *Advances in Neural Information Processing Systems*, 32.

Main Result



Theorem

$$\frac{1}{T_R^\star(\boldsymbol{\theta})} \leq \liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}[\tau_\pi]}{\log(1/\delta)} \leq \limsup_{\eta \downarrow 0} \limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{T_R^\star(\boldsymbol{\theta})}$$