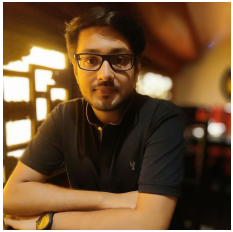
Optimal Best Arm Identification with Fixed Confidence in Restless Bandits

P. N. Karthik

Institute of Data Science
National University of Singapore
November 20, 2023

Joint Work With

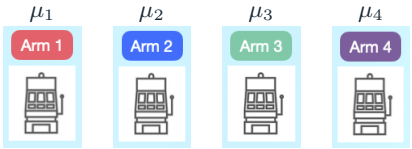
Vincent Tan
NUS, Singapore



Arpan Mukherjee
RPI, New York

Ali Tajer
RPI, New York

Preliminaries: Fixed-Confidence BAI

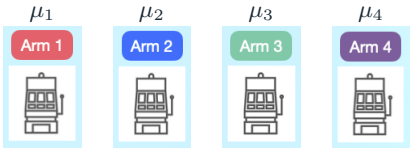


$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$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

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

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

Preliminaries: Fixed-Confidence BAI

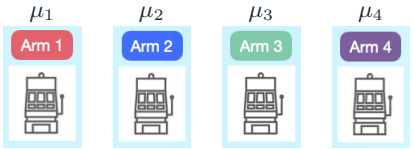


$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$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

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

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

Preliminaries: Fixed-Confidence BAI

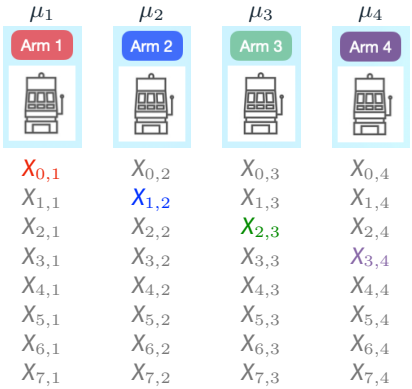


$X_{0,1}$	$X_{0,2}$	$X_{0,3}$	$X_{0,4}$
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$
$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$
$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$
$X_{7,1}$	$X_{7,2}$	$X_{7,3}$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

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

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

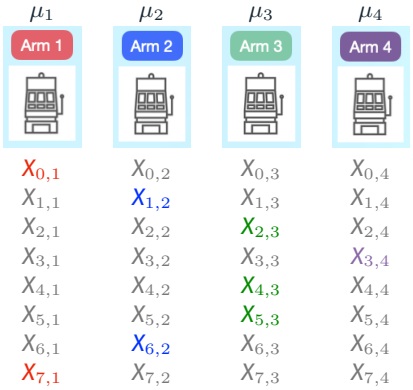
Preliminaries: Fixed-Confidence BAI



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	$X_{3,4}$

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

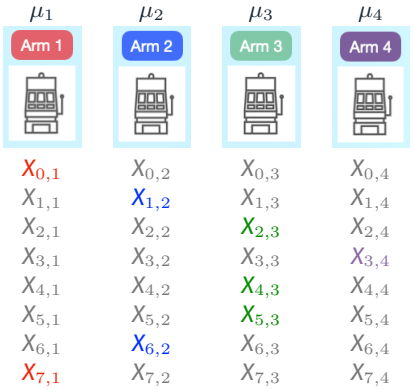
Preliminaries: Fixed-Confidence BAI



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	$X_{7,1}$

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

Preliminaries: Fixed-Confidence BAI

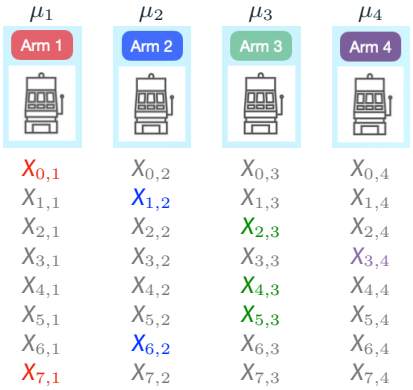


Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$
2	3	$X_{2,3}$
3	4	$X_{3,4}$
4	3	$X_{4,3}$
5	3	$X_{5,4}$
6	2	$X_{6,2}$
7	1	$X_{7,1}$

$$A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$$

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

Preliminaries: Fixed-Confidence BAI



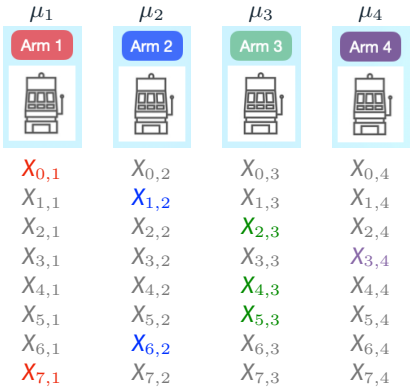
Time (n)	Arm (A_n)	Observation ($\bar{X}_n$)
0	1	$X_{0,1}$
1	2	$X_{1,2}$
2	3	$X_{2,3}$
3	4	$X_{3,4}$
4	3	$X_{4,3}$
5	3	$X_{5,4}$
6	2	$X_{6,2}$
7	1	$X_{7,1}$

$$A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$$

τ – stopping time

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

Preliminaries: Fixed-Confidence BAI



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	$X_{7,1}$

$A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$

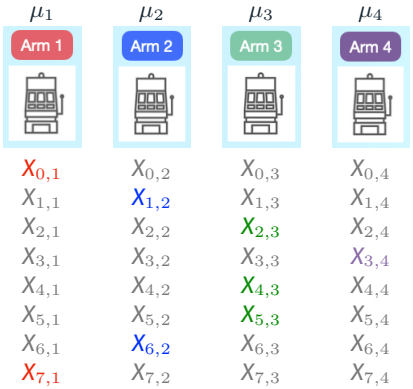
τ – stopping time

a – guess of the best arm

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$

$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

Preliminaries: Fixed-Confidence BAI



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	$X_{7,1}$

$A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$

τ – stopping time

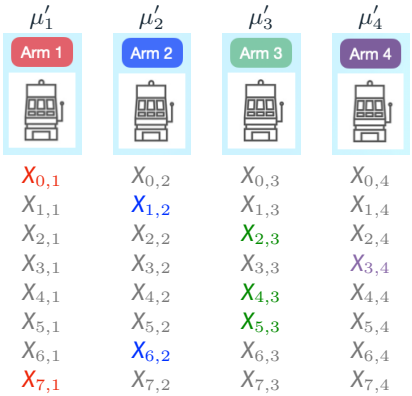
a – guess of the best arm

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4)$$
$$a^\star(\boldsymbol{\mu}) = \arg\max_a \mu_a$$

Given $\delta \in (0,1)$:

$$\mathbb{P}_{\boldsymbol{\mu}}(\tau < +\infty) = 1, \quad \mathbb{P}_{\boldsymbol{\mu}}(a = a^\star(\boldsymbol{\mu})) \geq 1 - \delta$$

Preliminaries: Fixed-Confidence BAI



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	$X_{7,1}$

$A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$

τ – stopping time

a – guess of the best arm

$$\boldsymbol{\mu'} = (\mu'_1, \ldots, \mu'_4)$$
$$a^\star(\boldsymbol{\mu'}) = \arg\max_a \mu'_a$$

Given $\delta \in (0,1)$:

$$\mathbb{P}_{\boldsymbol{\mu'}}(\tau < +\infty) = 1, \quad \mathbb{P}_{\boldsymbol{\mu'}}(a = a^\star(\boldsymbol{\mu'})) \geq 1 - \delta$$

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping
- $\pi = \{\pi_n\}_{n \geq 0}$

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping
- $\pi = \{\pi_n\}_{n \geq 0}$
- $\tau(\pi)$ – stopping time

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping
- $\pi = \{\pi_n\}_{n \geq 0}$
- $\tau(\pi)$ – stopping time
- $a(\pi)$ – guess of the best arm

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping
- $\pi = \{\pi_n\}_{n \geq 0}$
- $\tau(\pi)$ – stopping time
- $a(\pi)$ – guess of the best arm
- Given $\delta \in (0, 1)$, our interest is in the class

$$\Pi(\delta) := \left\{ \pi : \mathbb{P}_{\boldsymbol{\mu}'}(\tau(\pi) < +\infty) = 1, \quad \mathbb{P}_{\boldsymbol{\mu}'}(a(\pi) = a^\star(\boldsymbol{\mu}')) \geq 1 - \delta \quad \forall \boldsymbol{\mu}' \right\}$$

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping
- $\pi = \{\pi_n\}_{n \geq 0}$
- $\tau(\pi)$ – stopping time
- $a(\pi)$ – guess of the best arm
- Given $\delta \in (0, 1)$, our interest is in the class

$$\Pi(\delta) := \left\{ \pi : \mathbb{P}_{\boldsymbol{\mu}'}(\tau(\pi) < +\infty) = 1, \quad \mathbb{P}_{\boldsymbol{\mu}'}(a(\pi) = a^\star(\boldsymbol{\mu}')) \geq 1 - \delta \quad \forall \boldsymbol{\mu}' \right\}$$

Preliminaries: Fixed-Confidence BAI

- At time n: π_n – a rule for choosing $A_n \leftarrow (A_{0:n-1}, \bar{X}_{0:n-1})$ or stopping
- $\pi = \{\pi_n\}_{n \geq 0}$
- $\tau(\pi)$ – stopping time
- $a(\pi)$ – guess of the best arm
- Given $\delta \in (0,1)$, our interest is in the class

$$\Pi(\delta) := \left\{ \pi : \mathbb{P}_{\boldsymbol{\mu}'}(\tau(\pi) < +\infty) = 1, \quad \mathbb{P}_{\boldsymbol{\mu}'}(a(\pi) = a^\star(\boldsymbol{\mu}')) \geq 1 - \delta \quad \forall \boldsymbol{\mu}' \right\}$$

How does $\inf_{\pi \in \Pi(\delta)} \mathbb{E}_{\boldsymbol{\mu}}[\tau(\pi)]$ grow as $\delta \downarrow 0$ for any given $\boldsymbol{\mu}$?

Preliminaries: Fixed-Confidence BAI

Theorem

Garivier and Kaufmann (2016)
Assuming there are K arms and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$,

$$\lim_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau(\pi)]}{\log(1/\delta)} \geq \frac{1}{T^\star(\boldsymbol{\mu})},$$

where the constant $T^\star(\boldsymbol{\mu})$ is defined as

$$T^\star(\boldsymbol{\mu}) := \sup_{\nu \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{ALT}(\boldsymbol{\mu})} \sum_{a=1}^{K} \nu(a) \, \mathrm{KL}(\mu_a, \lambda_a)$$

$$\mathrm{ALT}(\boldsymbol{\mu}) := \left\{ \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K) : a^\star(\boldsymbol{\lambda}) \neq a^\star(\boldsymbol{\mu}) \right\}$$

Preliminaries: Fixed-Confidence BAI

$$T^\star(\boldsymbol{\mu}) = \sup_{\nu \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu})} \sum_{a=1}^{K} \nu(a)\, \mathrm{KL}(\mu_a, \lambda_a)$$

$$\psi(\nu, \boldsymbol{\mu}) = \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu})} \sum_{a=1}^{K} \nu(a)\, \mathrm{KL}(\mu_a, \lambda_a), \quad \nu \in \Sigma_K,\ \boldsymbol{\mu} \in \mathbb{R}^K, \qquad \nu^\star_{\boldsymbol{\mu}} = \arg\sup_{\nu} \psi(\nu, \boldsymbol{\mu})$$

Parameter Estimation

1

$$\widehat{\boldsymbol{\mu}}(n) = [\widehat{\mu}_1(n), \ldots, \widehat{\mu}_K(n)]^\top$$

Optimal Distribution

2

$$\nu^\star_{\widehat{\boldsymbol{\mu}}(n)} = \arg\sup_{\nu} \psi(\nu, \widehat{\boldsymbol{\mu}}(n))$$

Certainty Equivalence

3

$$A_{n+1} \sim \eta\, \nu^{\text{unif}} + (1-\eta)\, \nu^\star_{\widehat{\boldsymbol{\mu}}(n)}$$

Key ideas: $\widehat{\boldsymbol{\mu}}(n) \overset{n\to\infty}{\longrightarrow} \boldsymbol{\mu}$, $\quad \nu^\star_{\widehat{\boldsymbol{\mu}}(n)}(a) \overset{n\to\infty}{\longrightarrow} \nu^\star_{\boldsymbol{\mu}}(a)$, $\quad \dfrac{N_a(n)}{n} \overset{n\to\infty}{\longrightarrow} \nu^\star_{\boldsymbol{\mu}}(a)$ a.s. $\forall a$

Preliminaries: Fixed-Confidence BAI

Parameter Estimation

1

$$\widehat{\boldsymbol{\mu}}(n) = [\widehat{\mu}_1(n), \ldots, \widehat{\mu}_K(n)]^\top$$

Optimal Distribution

2

$$\nu^\star_{\widehat{\boldsymbol{\mu}}(n)} = \arg\sup_{\nu} \psi(\nu, \widehat{\boldsymbol{\mu}}(n))$$

Certainty Equivalence

3

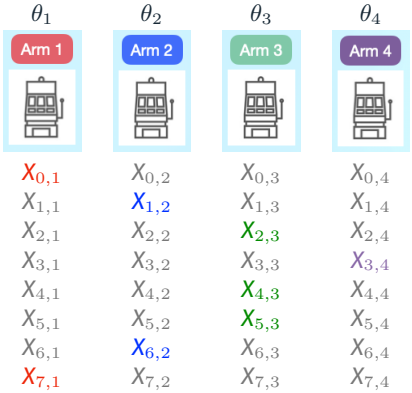
$$A_{n+1} \sim \eta\, \nu^{\text{unif}} + (1-\eta)\, \nu^\star_{\widehat{\boldsymbol{\mu}}(n)}$$

Theorem

For a threshold-based stopping rule using the GLLR along with the above arms selection rule,

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

Preliminaries: Fixed-Confidence BAI



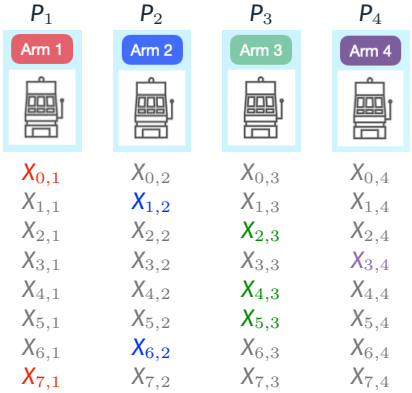
$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_4)$$

$$a^\star(\boldsymbol{\theta}) = \arg\max_a \theta_a$$

- Single-parameter exponential family (SPEF)
- $T^\star(\boldsymbol{\theta}) = \sup_{\nu} \inf_{\lambda \in \mathsf{ALT}(\boldsymbol{\theta})} \sum_{a=1}^{K} \nu(a)\, \mathrm{KL}(\theta_a, \lambda_a)$
- $\theta \iff \mu_\theta = \mathbb{E}_\theta[X]$
  (one-to-one correspondence)

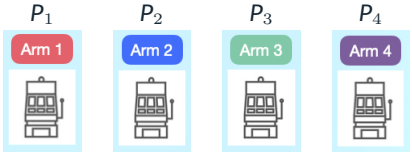
This talk: extensions to **restless Markovian** arms

BAI with Restless Markovian Arms



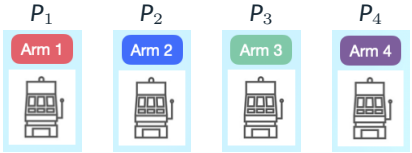
- Common, finite state space
- P_a is ergodic for each a
- μ_a – stationary distribution of P_a

BAI with Restless Markovian Arms



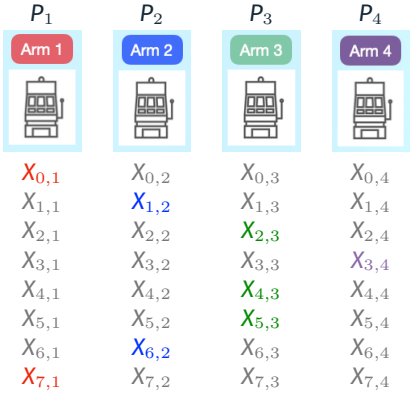
- Common, finite state space
- P_a is ergodic for each a
- μ_a – stationary distribution of P_a
- How to define "mean" of arm a?

BAI with Restless Markovian Arms



- Common, finite state space
- P_a is ergodic for each a
- μ_a – stationary distribution of P_a
- How to define "mean" of arm a?
- Mean $\Longleftrightarrow$ TPM?

BAI with Restless Markovian Arms



- Common, finite state space
- P_a is ergodic for each a
- μ_a – stationary distribution of P_a
- How to define "mean" of arm a?
- Mean $\Longleftrightarrow$ TPM?

Answer: SPEF for Markov chains!

Model: Exponential Family of TPMs

$\Theta \subset \mathbb{R}$ - known parameter space
P – irreducible TPM on $\mathcal{S}$ (generator)
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

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

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

Model: Exponential Family of TPMs

$\Theta \subset \mathbb{R}$ - known parameter space
P – irreducible TPM on $\mathcal{S}$ (generator)
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

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

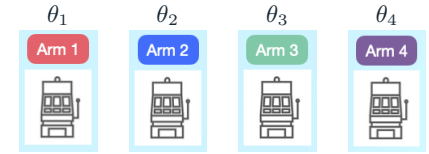
$$\theta \mapsto P_\theta(i,j) = \frac{v_\theta(j)}{\rho(\theta)\, v_\theta(i)} \times \tilde{P}_\theta(i,j), \quad i,j \in \mathcal{S}$$

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

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

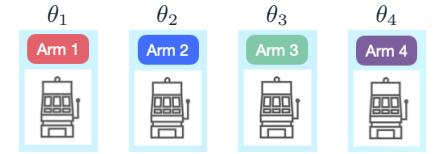
Analysis

Analysis



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

Analysis



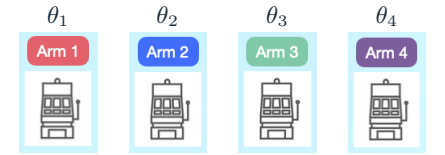
Arm delays

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

Analysis



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

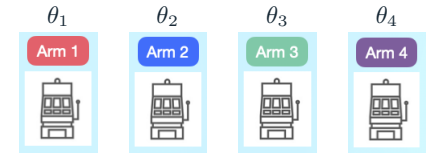
Arm delays

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				

Analysis



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

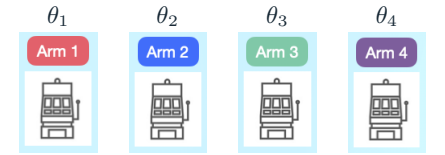
Arm delays

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	$X_{7,1}$			

Analysis



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

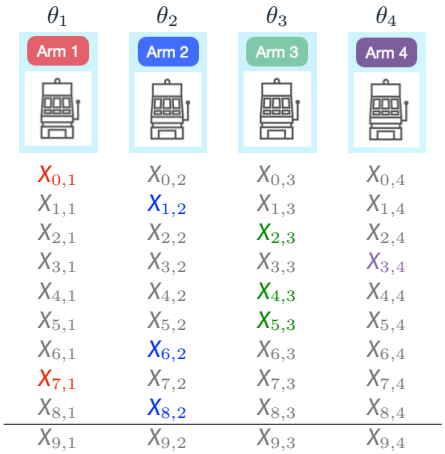
Arm delays

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	$X_{7,1}$	$X_{8,2}$		

Analysis



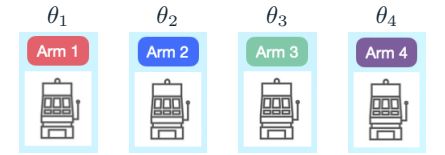
Arm delays

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	$X_{7,1}$	$X_{8,2}$	$X_{5,3}$	

Analysis



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

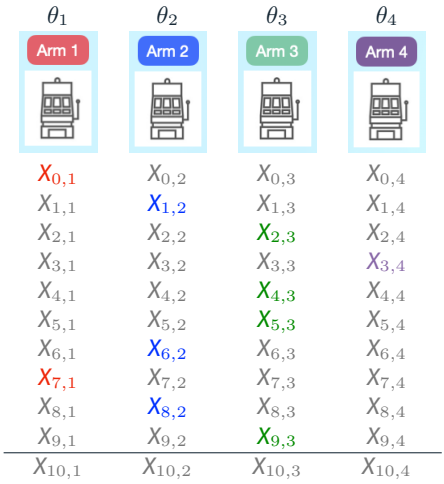
Arm delays

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

Analysis



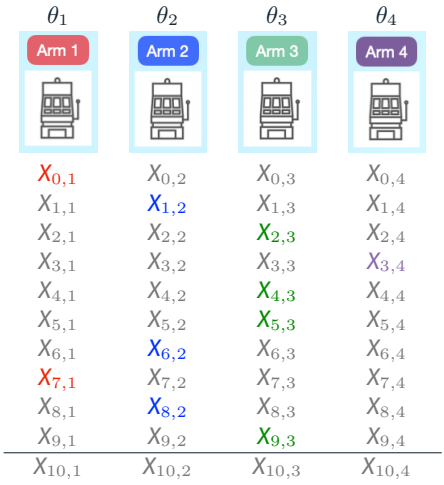
Arm delays

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				

Analysis



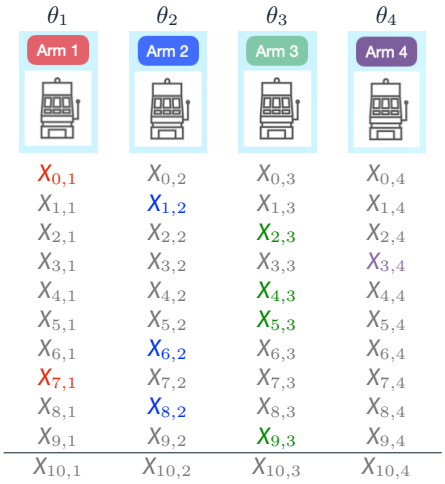
Arm delays

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	$X_{7,1}$			

Analysis



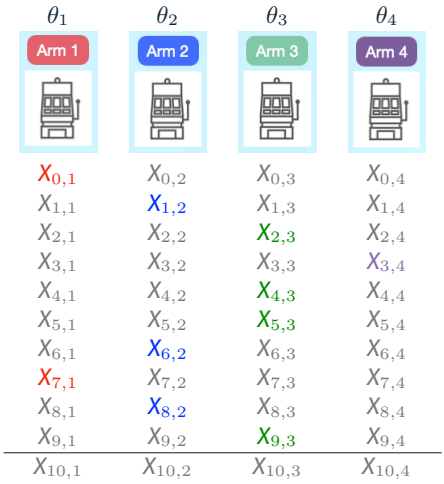
Arm delays

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	$X_{7,1}$	$X_{8,2}$		

Analysis



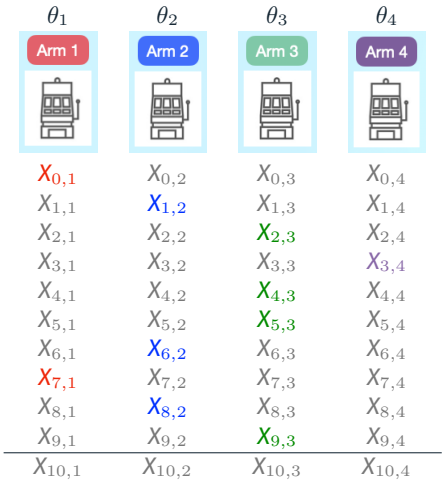
Arm delays

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	$X_{7,1}$	$X_{8,2}$	$X_{9,3}$	

Analysis



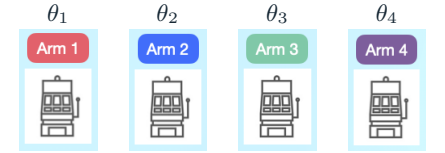
Arm delays

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

Analysis



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

Arm delays

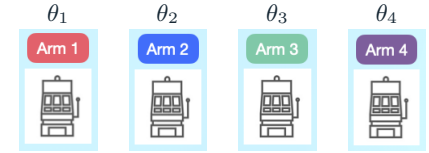
n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

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

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

Analysis



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

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

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

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

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

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

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

Markov Decision Process

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

Action space: $[K]$ finite

Arm delays

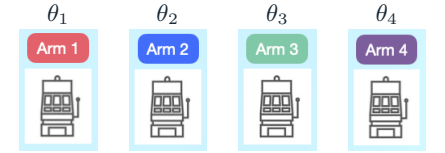
n	$d_1(n)$	$d_2(n)$	$d_3(n)$	$d_4(n)$
8	1	2	3	5
9	2	1	4	6
10	3	2	1	7

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

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

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

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

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

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	7

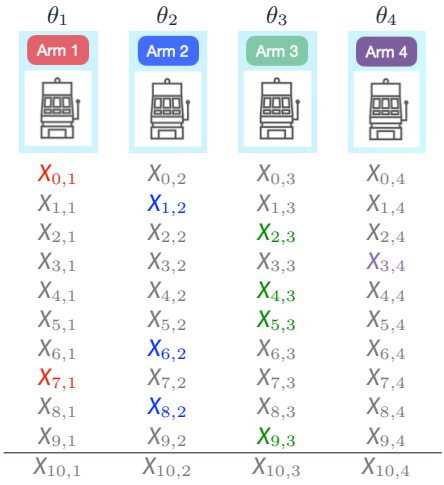
Last observed states

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

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

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

Flow Constraint



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

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

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

Flow Constraint

Under every policy π,

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

Constraint on Maximum Delay

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
- $\mathbb{S}_R = \{(\mathbf{d}, \mathbf{i}) \in \mathbb{S} : \max_a d_a \leq R\}, \qquad \mathbb{S}_{R,a} = \{(\mathbf{d}, \mathbf{i}) \in \mathbb{S}_R : d_a = R\}$ finite state space
- $\Pr_{\boldsymbol{\theta}} \longrightarrow \Pr_{\boldsymbol{\theta},R}$

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
- $\mathbb{S}_R = \{(\mathbf{d}, \mathbf{i}) \in \mathbb{S} : \max_a d_a \leq R\}, \qquad \mathbb{S}_{R,a} = \{(\mathbf{d}, \mathbf{i}) \in \mathbb{S}_R : d_a = R\}$ finite state space
- $\Pr_{\theta} \longrightarrow \Pr_{\theta,R}$

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

$R = 6$

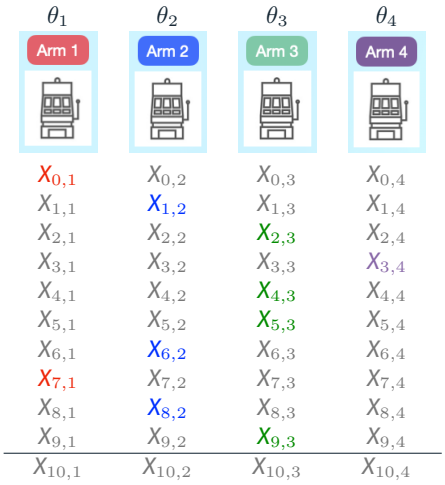
$(\mathbf{d}(9), \mathbf{i}(9)) \longrightarrow A_{10} = 4$ ✔

Flow Constraint + R-Max-Delay Constraint

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

$N(n, \mathbf{d}, \mathbf{i}, a) = N(n, \mathbf{d}, \mathbf{i}) \quad \text{if } d_a = R$

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:

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

Converse

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

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

Proposition

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

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

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

Converse – 1

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

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

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

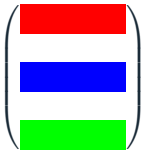
$$\liminf_{\delta \downarrow 0} \inf_{\pi \in \Pi(\delta)} \frac{\mathbb{E}[\tau_\pi]}{\log(1/\delta)} \geq \frac{1}{T_R^\star(\boldsymbol{\theta})}$$

Converse – 2

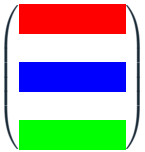
$$T_R^\star(\boldsymbol{\theta}) = \sup_{\nu \in \Sigma_R(\boldsymbol{\theta})} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\theta})} \sum_{(\mathbf{d},\mathbf{i}) \in \mathbb{S}_R} \sum_{a=1}^{K} \nu(\mathbf{d},\mathbf{i},a)\, D_{\text{KL}}(\Pr_{\boldsymbol{\theta},R}(\cdot|\mathbf{d},\mathbf{i},a)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

Converse – 2

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



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



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

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

Converse – 3

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

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

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

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

$$\nu(\mathbf{d},\mathbf{i},a) = \sum_{a'=1}^{K} \nu(\mathbf{d},\mathbf{i},a') \quad \text{if } d_a = R,\ a \in [K] \Bigg\}$$

Converse – 4

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

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

Lemma

Converse – 4

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

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

Lemma

- ψ *is continuous*

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

Converse – 4

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

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

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

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

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

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

Comparison

Restless arms

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

Independent observations from arms

Garivier and Kaufmann (2016)

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

Achievability

Uniform Policy and Ergodicity

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

Uniform Policy and Ergodicity

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

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

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

Uniform Policy and Ergodicity

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

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

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

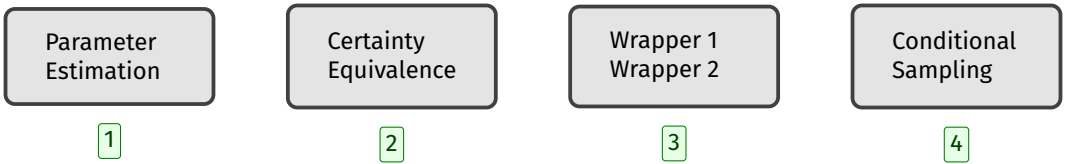
Lemma

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

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

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

Achievability – Arm Selection Rule



Arm Selection Rule – 1

Parameter Estimation

Certainty Equivalence

Wrapper 1
Wrapper 2

Conditional Sampling

1

2

3

4

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

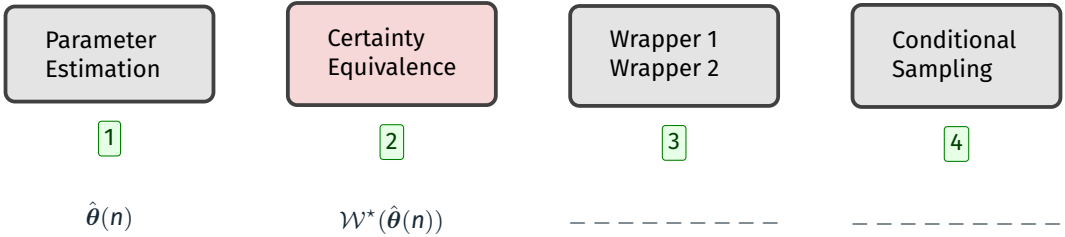
$$N_a(n) = \sum_{t=0}^{n} \mathbf{1}_{\{A_t = a\}}$$

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

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

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

Arm Selection Rule – 2



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

Arm Selection Rule – 3

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

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

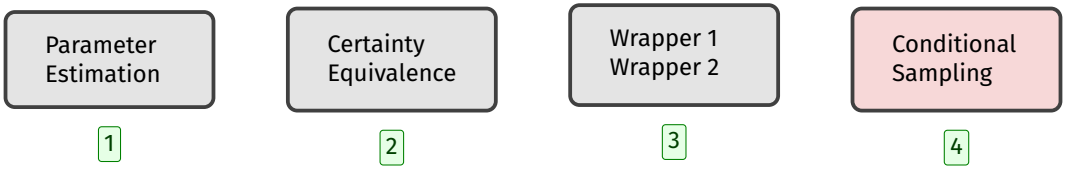
Wrapper 1

$$\pi^{\eta}_{\hat{\boldsymbol{\theta}}(n)}(a|\mathbf{d},\mathbf{i}) = \frac{\eta\, \nu^{\mathsf{unif}}_{\hat{\boldsymbol{\theta}}(n)}(\mathbf{d},\mathbf{i},a) + (1-\eta)\, \nu_n^\star(\mathbf{d},\mathbf{i},a)}{\eta\, \mu^{\mathsf{unif}}_{\hat{\boldsymbol{\theta}}(n)}(\mathbf{d},\mathbf{i}) + (1-\eta) \sum_{a'=1}^{K} \nu_n^\star(\mathbf{d},\mathbf{i},a')}$$

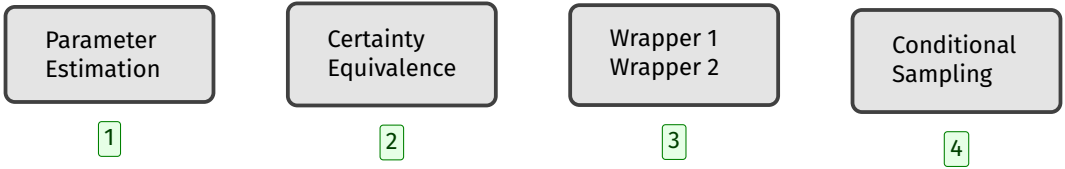
Wrapper 2

$$\pi_n = \varepsilon_n \pi^{\mathsf{unif}} + (1-\varepsilon_n)\, \pi^{\eta}_{\hat{\boldsymbol{\theta}}(n-1)} \quad \forall n \geq K$$

Arm Selection Rule – 4



Achievability – Arm Selection Rule



Lemma

Achievability – Arm Selection Rule

Parameter Estimation	Certainty Equivalence	Wrapper 1 Wrapper 2	Conditional Sampling
1	2	3	4
$\hat{\boldsymbol{\eta}}(n) \Longleftrightarrow \hat{\boldsymbol{\theta}}(n)$	$\mathcal{W}^\star(\hat{\boldsymbol{\theta}}(n))$	η, ε_n-mixtures	$A_n \sim \pi_n(\cdot \mid \mathbf{d}(n), \mathbf{i}(n))$

Lemma

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

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

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:

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

Achievability – Stopping Rule

Empirical transition matrix:

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

Test statistic

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

Achievability – Stopping Rule

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

Threshold:

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

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

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

- Upper bound in expectation

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

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

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 $a \in [K]$, $N(\mathbf{n}, \mathbf{d}, \mathbf{i}, a) = 0$ for all $(\mathbf{d}, \mathbf{i}, a) \in \mathbb{S}_R \times [K]$, $\texttt{stop} = 0$

for $n < K$ **do**
  Select arm $A_n = n + 1$.
end for
while $\texttt{stop} == 0$ **do**
  Update $(\mathbf{d}(n), \mathbf{i}(n))$; 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 $A_n \sim \pi_n(\cdot | \mathbf{d}(n), \mathbf{i}(n))$
    $n \leftarrow n + 1$.
  end if
end while
return $\hat{a}$.

Main Result

Parameter Estimation	Certainty Equivalence	Wrapper 1 Wrapper 2	Conditional Sampling
1	2	3	4
$\hat{\boldsymbol{\eta}}(n) \Longleftrightarrow \hat{\boldsymbol{\theta}}(n)$	$\mathcal{W}^\star(\hat{\boldsymbol{\theta}}(n))$	η, ε_n-mixtures	$A_n \sim \pi_n(\cdot \mid \mathbf{d}(n), \mathbf{i}(n))$

Theorem

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

Removing the Max-Delay Constraint

Removing the Max-Delay Constraint

Lemma (Montonicity)

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

Removing the Max-Delay Constraint

Lemma (Montonicity)

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

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

Removing the Max-Delay Constraint

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

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

Removing the Max-Delay Constraint

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$\Big\}$

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

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

Removing the Max-Delay Constraint

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

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

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

Independent observations from arms ✔

Restless arms – OPEN

Removing the Max-Delay Constraint

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

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

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

Independent observations from arms ✔

Restless arms – OPEN

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

Main Result

Parameter Estimation	Certainty Equivalence	Wrapper 1 Wrapper 2	Conditional Sampling
1	2	3	4
$\hat{\boldsymbol{\eta}}(n) \Longleftrightarrow \hat{\boldsymbol{\theta}}(n)$	$\mathcal{W}^\star(\hat{\boldsymbol{\theta}}(n))$	η, ε_n-mixtures	$A_n \sim \pi_n(\cdot \mid \mathbf{d}(n), \mathbf{i}(n))$

Theorem

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

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.