Visual Search with a Trembling Hand

An Analysis of Odd Arm Identification in Restless Multi-armed Bandits

P. N. Karthik
periyapatna@iisc.ac.in

Networks Seminar Presentation
RBCCPS, IISc
March 10, 2020

Motivation

A Visual Search Experiment

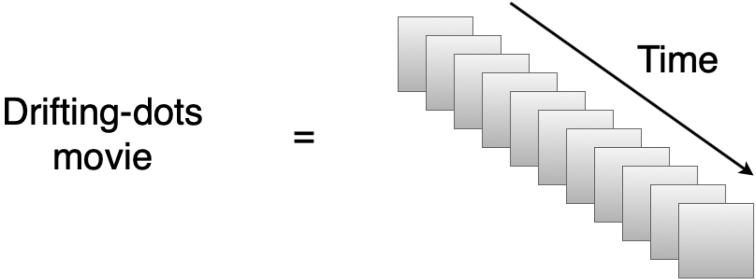
- A total of 8 drifting-dots moving images
- The angle of movement (drift) in the of the movies is different
- **Goal: Identify the "odd" movie as quickly as possible**

A Visual Search Experiment

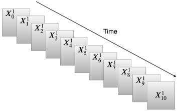
- A total of 8 drifting-dots moving images
- The angle of movement (drift) in the of the movies is different
- **Goal: Identify the “odd” movie as quickly as possible**

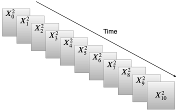
What is the relation between
(a) the time taken to identify the odd movie, and
(b) the “closeness” of the odd and the non-odd movies?

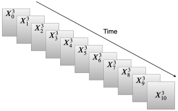
Demystifying the Experiment



Demystifying the Experiment: 3 Movies





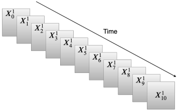


$\{X_t^1 : t = 0, 1, 2, \ldots\}$ $\{X_t^2 : t = 0, 1, 2, \ldots\}$ $\{X_t^3 : t = 0, 1, 2, \ldots\}$

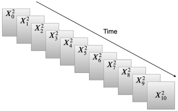
Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$			

Delay: Time since last observation **LOS: Last Observed State**

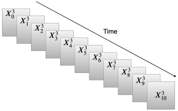
Demystifying the Experiment: 3 Movies



$\{X_t^1 : t = 0, 1, 2, \ldots\}$



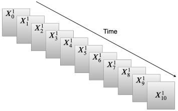
$\{X_t^2 : t = 0, 1, 2, \ldots\}$



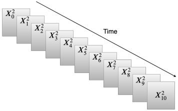
$\{X_t^3 : t = 0, 1, 2, \ldots\}$

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$			

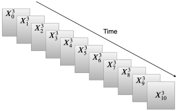
Demystifying the Experiment: 3 Movies



$\{X_t^1 : t = 0, 1, 2, \ldots\}$



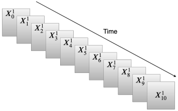
$\{X_t^2 : t = 0, 1, 2, \ldots\}$



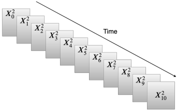
$\{X_t^3 : t = 0, 1, 2, \ldots\}$

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$			

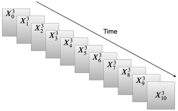
Demystifying the Experiment: 3 Movies



$$\{X_t^1 : t = 0, 1, 2, \ldots\}$$



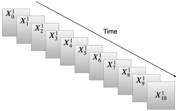
$$\{X_t^2 : t = 0, 1, 2, \ldots\}$$



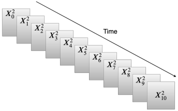
$$\{X_t^3 : t = 0, 1, 2, \ldots\}$$

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$	$d_1(t) = 2$	$d_2(t) = 3$	$d_3(t) = 1$
	$i_1(t) = i_i(t-1)$	$i_2(t) = i_2(t-1)$	$i_3(t) = X_6^3$
$t = 7$			

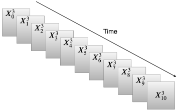
Demystifying the Experiment: 3 Movies



$\{X_t^1 : t = 0, 1, 2, \ldots\}$



$\{X_t^2 : t = 0, 1, 2, \ldots\}$

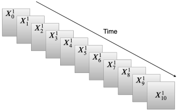


$\{X_t^3 : t = 0, 1, 2, \ldots\}$

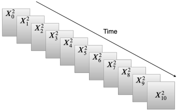
Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$	$d_1(t) = 2$	$d_2(t) = 3$	$d_3(t) = 1$
	$i_1(t) = i_i(t-1)$	$i_2(t) = i_2(t-1)$	$i_3(t) = X_6^3$
$t = 7$	$d_1(t) = 1$	$d_2(t) = 4$	$d_3(t) = 2$
	$i_1(t) = X_7^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 8$			

Delay: Time since last observation **LOS: Last Observed State**

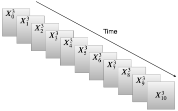
Demystifying the Experiment: 3 Movies



$\{X_t^1 : t = 0, 1, 2, \ldots\}$



$\{X_t^2 : t = 0, 1, 2, \ldots\}$



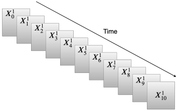
$\{X_t^3 : t = 0, 1, 2, \ldots\}$

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$	$d_1(t) = 2$	$d_2(t) = 3$	$d_3(t) = 1$
	$i_1(t) = i_i(t-1)$	$i_2(t) = i_2(t-1)$	$i_3(t) = X_6^3$
$t = 7$	$d_1(t) = 1$	$d_2(t) = 4$	$d_3(t) = 2$
	$i_1(t) = X_7^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 8$	$d_1(t) = 2$	$d_2(t) = 1$	$d_3(t) = 3$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_8^2$	$i_3(t) = i_3(t-1)$
$t = 9$			

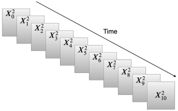
Delay: Time since last observation

LOS: Last Observed State

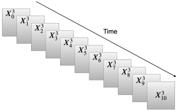
Demystifying the Experiment: 3 Movies



$\{X_t^1 : t = 0, 1, 2, \ldots\}$



$\{X_t^2 : t = 0, 1, 2, \ldots\}$



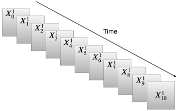
$\{X_t^3 : t = 0, 1, 2, \ldots\}$

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$	$d_1(t) = 2$	$d_2(t) = 3$	$d_3(t) = 1$
	$i_1(t) = i_i(t-1)$	$i_2(t) = i_2(t-1)$	$i_3(t) = X_6^3$
$t = 7$	$d_1(t) = 1$	$d_2(t) = 4$	$d_3(t) = 2$
	$i_1(t) = X_7^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 8$	$d_1(t) = 2$	$d_2(t) = 1$	$d_3(t) = 3$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_8^2$	$i_3(t) = i_3(t-1)$
$t = 9$	$d_1(t) = 3$	$d_2(t) = 1$	$d_3(t) = 4$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_9^2$	$i_3(t) = i_3(t-1)$
$t = 10$			

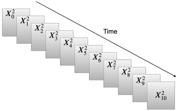
Delay: Time since last observation

LOS: Last Observed State

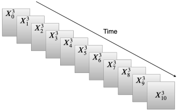
Demystifying the Experiment: 3 Movies



$\{X_t^1 : t = 0, 1, 2, \ldots\}$



$\{X_t^2 : t = 0, 1, 2, \ldots\}$



$\{X_t^3 : t = 0, 1, 2, \ldots\}$

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$	$d_1(t) = 2$	$d_2(t) = 3$	$d_3(t) = 1$
	$i_1(t) = i_i(t-1)$	$i_2(t) = i_2(t-1)$	$i_3(t) = X_6^3$
$t = 7$	$d_1(t) = 1$	$d_2(t) = 4$	$d_3(t) = 2$
	$i_1(t) = X_7^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 8$	$d_1(t) = 2$	$d_2(t) = 1$	$d_3(t) = 3$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_8^2$	$i_3(t) = i_3(t-1)$
$t = 9$	$d_1(t) = 3$	$d_2(t) = 1$	$d_3(t) = 4$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_9^2$	$i_3(t) = i_3(t-1)$
$t = 10$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 5$
	$i_1(t) = X_{10}^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$

Delay: Time since last observation

LOS: Last Observed State

Odd Arm Identification

Visual Search Experiment vs Multi-armed Bandits

Visual Search Experiment	**Multi-armed Bandits**
Movies	Arms
X_t^a: neuronal response to movie a at time t	X_t^a: observation from arm a at time t
Movie **paused** when not observed	Arms **rested**
Movie **continues** when not observed	Arms **restless**
Finding the odd movie	Finding the odd arm

Visual Search Experiment vs Multi-armed Bandits

Visual Search Experiment	Multi-armed Bandits
Movies	Arms
X_t^a: neuronal response to movie a at time t	X_t^a: observation from arm a at time t
Movie **paused** when not observed	Arms **rested**
Movie **continues** when not observed	Arms **restless**
Finding the odd movie	Finding the odd arm

Odd arm identification with Markov observations:

- $K \geq 3$ arms

Visual Search Experiment vs Multi-armed Bandits

Visual Search Experiment	Multi-armed Bandits
Movies	Arms
X_t^a: neuronal response to movie a at time t	X_t^a: observation from arm a at time t
Movie **paused** when not observed	Arms **rested**
Movie **continues** when not observed	Arms **restless**
Finding the odd movie	Finding the odd arm

Odd arm identification with Markov observations:

- $K \geq 3$ arms
- $\{X_t^a : t = 0, 1, 2, \ldots\}$: Markov process of arm a

Visual Search Experiment vs Multi-armed Bandits

Visual Search Experiment	**Multi-armed Bandits**
Movies	Arms
X_t^a: neuronal response to movie a at time t	X_t^a: observation from arm a at time t
Movie **paused** when not observed	Arms **rested**
Movie **continues** when not observed	Arms **restless**
Finding the odd movie	Finding the odd arm

Odd arm identification with Markov observations:

- $K \geq 3$ arms
- $\{X_t^a : t = 0, 1, 2, \ldots\}$: Markov process of arm a
- Finite state space $\mathcal{S}$

Visual Search Experiment vs Multi-armed Bandits

Visual Search Experiment	**Multi-armed Bandits**
Movies	Arms
X_t^a: neuronal response to movie a at time t	X_t^a: observation from arm a at time t
Movie **paused** when not observed	Arms **rested**
Movie **continues** when not observed	Arms **restless**
Finding the odd movie	Finding the odd arm

Odd arm identification with Markov observations:

- $K \geq 3$ arms
- $\{X_t^a : t = 0, 1, 2, \ldots\}$: Markov process of arm a
- Finite state space $\mathcal{S}$
- TPM of one of the arms is P_1 (odd arm)
- TPM of rest of the arms is P_2, where $P_2 \neq P_1$

Visual Search Experiment vs Multi-armed Bandits

Visual Search Experiment	Multi-armed Bandits
Movies	Arms
X_t^a: neuronal response to movie a at time t	X_t^a: observation from arm a at time t
Movie **paused** when not observed	Arms **rested**
Movie **continues** when not observed	Arms **restless**
Finding the odd movie	Finding the odd arm

Odd arm identification with Markov observations:

- $K \geq 3$ arms
- $\{X_t^a : t = 0, 1, 2, \ldots\}$: Markov process of arm a
- Finite state space $\mathcal{S}$
- TPM of one of the arms is P_1 (odd arm)
- TPM of rest of the arms is P_2, where $P_2 \neq P_1$
- P_1 and P_2 may be known or unknown

Active Sequential Hypothesis Testing

When P_1 and P_2 are known: K simple hypotheses

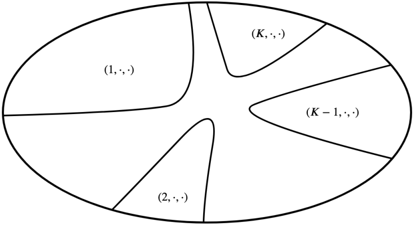
$$
\begin{aligned}
\mathcal{H}_1 &: \text{arm 1 is the odd arm} \\
\mathcal{H}_2 &: \text{arm 2 is the odd arm} \\
&\vdots \\
\mathcal{H}_K &: \text{arm } K \text{ is the odd arm}
\end{aligned}
$$

Active Sequential Hypothesis Testing

When P_1 and P_2 are known: K simple hypotheses

$$\mathcal{H}_1 : \text{arm 1 is the odd arm}$$
$$\mathcal{H}_2 : \text{arm 2 is the odd arm}$$
$$\vdots$$
$$\mathcal{H}_K : \text{arm } K \text{ is the odd arm}$$

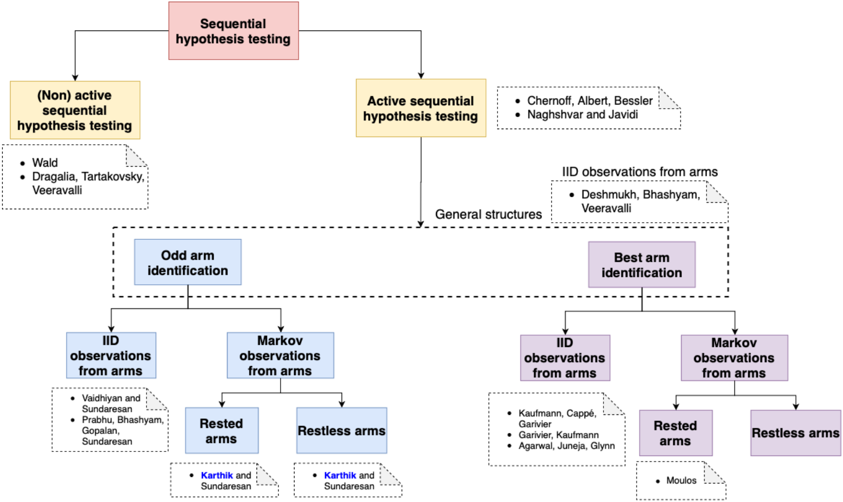
When neither P_1 nor P_2 is known: K composite hypotheses



$$\mathcal{H}_1 : (1, \cdot, \cdot)$$
$$\mathcal{H}_2 : (2, \cdot, \cdot)$$
$$\vdots$$
$$\mathcal{H}_K : (K, \cdot, \cdot)$$

Notation : (odd arm, P_1, P_2)

The Big Picture



Odd Arm Identification in Restless Multi-armed Bandits [a]

[a] P. N. Karthik and R. Sundaresan, "Detecting an Odd Restless Markov Arm with a Trembling Hand". Submitted to ISIT 2020.

Preparing The Ground

- Assume P_1 and P_2 are known ($P_2 \neq P_1$)
- Unknown: the odd arm

Preparing The Ground

- Assume P_1 and P_2 are known ($P_2 \neq P_1$)
- Unknown: the odd arm

Two quantities of interest:

(a) (Average) time taken to identify the odd arm

(b) Probability of error

Preparing The Ground

- Assume P_1 and P_2 are known ($P_2 \neq P_1$)
- Unknown: the odd arm

Two quantities of interest:

(a) (Average) time taken to identify the odd arm

(b) Probability of error

This talk:

- **Fixed confidence setting**
- **Asymptotics as (b)** $\to 0$

Preparing The Ground

- Assume P_1 and P_2 are known ($P_2 \neq P_1$)
- Unknown: the odd arm

Two quantities of interest:

(a) (Average) time taken to identify the odd arm

(b) Probability of error

This talk:

- **Fixed confidence setting**
- **Asymptotics as (b)** $\to 0$

Main results:

- **An asymptotic lower bound on (a)**
- **A sequence of strategies that achieve the lower bound asymptotically**

Notations and Preliminaries

A Few Notations

A_t: arm selected at time t
$X_t^{A_t}$: observation (Markov state) from arm A_t at time t

A Few Notations

A_t: arm selected at time t
$X_t^{A_t}$: observation (Markov state) from arm A_t at time t

Policy π: is a mapping

$$(A_0, X_0^{A_0}, A_1, X_1^{A_1}, \ldots, A_t, X_t^{A_t}) \mapsto A_{t+1}$$

$\tau(\pi)$: stopping time of policy π
$\theta(\pi)$: odd arm output by policy π

A Few Notations

A_t: arm selected at time t
$X_t^{A_t}$: observation (Markov state) from arm A_t at time t

Policy π: is a mapping

$$(A_0, X_0^{A_0}, A_1, X_1^{A_1}, \ldots, A_t, X_t^{A_t}) \mapsto A_{t+1}$$

$\tau(\pi)$: stopping time of policy π
$\theta(\pi)$: odd arm output by policy π

$P_h(\cdot)$: probabilities computed when h is the odd arm

Given a probability of error $\epsilon > 0$, our interest is in

$$\Pi(\epsilon) = \{\pi : P_h(\theta(\pi) \neq h) \leq \epsilon \text{ for all } h\}$$

A Few Notations

Given a probability of error $\epsilon > 0$, our interest is in

$$\Pi(\epsilon) = \{\pi : P_h(\theta(\pi) \neq h) \leq \epsilon \text{ for all } h\}$$

A Few Notations

Given a probability of error $\epsilon > 0$, our interest is in

$$\Pi(\epsilon) = \{\pi : P_h(\theta(\pi) \neq h) \leq \epsilon \text{ for all } h\}$$

From prior works, we expect

$$\inf_{\pi \in \Pi(\epsilon)} \mathbb{E}_h[\tau(\pi)] = \Theta\left(\log \frac{1}{\epsilon}\right)$$

A Few Notations

Given a probability of error $\epsilon > 0$, our interest is in

$$\Pi(\epsilon) = \{\pi : P_h(\theta(\pi) \neq h) \leq \epsilon \text{ for all } h\}$$

From prior works, we expect

$$\inf_{\pi \in \Pi(\epsilon)} \mathbb{E}_h[\tau(\pi)] = \Theta\left(\log \frac{1}{\epsilon}\right)$$

Our objective: to characterise

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{\mathbb{E}_h[\tau(\pi)]}{\log(1/\epsilon)}$$

for the setting of restless Markov arms

The Notion of a Trembling Hand

A_t: arm selected at time t
$X_t^{A_t}$: observation (Markov state) from arm A_t at time t
Policy π is a mapping

$$(A_0, X_0^{A_0}, A_1, X_1^{A_1}, \ldots, A_t, X_t^{A_t}) \mapsto A_{t+1}$$

The Notion of a Trembling Hand

A_t: arm selected at time t

$X_t^{A_t}$: observation (Markov state) from arm A_t at time t

Policy π is a mapping

$$(A_0, X_0^{A_0}, A_1, X_1^{A_1}, \ldots, A_t, X_t^{A_t}) \mapsto A_{t+1}$$

Trembling hand:

Often, what you select is B_t, but what ends up getting selected is A_t, where

$$A_t = \begin{cases} B_t, & \text{w.p. } 1-\eta, \\ \text{unif. randomly chosen arm}, & \text{w.p. } \eta \end{cases}$$

for some small $\eta > 0$

The Notion of a Trembling Hand

A_t: arm selected at time t

$X_t^{A_t}$: observation (Markov state) from arm A_t at time t

Policy π is a mapping

$$(A_0, X_0^{A_0}, A_1, X_1^{A_1}, \ldots, A_t, X_t^{A_t}) \mapsto A_{t+1}$$

Trembling hand:

Often, what you select is B_t, but what ends up getting selected is A_t, where

$$A_t = \begin{cases} B_t, & \text{w.p. } 1-\eta, \\ \text{unif. randomly chosen arm}, & \text{w.p. } \eta \end{cases}$$

for some small $\eta > 0$

Updated policy π:

$$(B_0, A_0, X_0^{A_0}, B_1, A_1, X_1^{A_1}, \ldots, B_t, A_t, X_t^{A_t}) \mapsto B_{t+1}$$

A New Notion of State

Time	**Delay** of movie 1	**Delay** of movie 2	**Delay** of movie 3
	LOS of movie 1	**LOS** of movie 2	**LOS** of movie 3
$t=3$	$d_1(t)=3$	$d_2(t)=2$	$d_3(t)=1$
	$i_1(t)=X_0^1$	$i_2(t)=X_1^2$	$i_3(t)=X_2^3$
$t=4$	$d_1(t)=4$	$d_2(t)=1$	$d_3(t)=2$
	$i_1(t)=i_1(t-1)$	$i_2(t)=X_3^2$	$i_3(t)=i_3(t-1)$
$t=5$	$d_1(t)=1$	$d_2(t)=2$	$d_3(t)=3$
	$i_1(t)=X_5^1$	$i_2(t)=i_2(t-1)$	$i_3(t)=i_3(t-1)$
$t=6$	$d_1(t)=2$	$d_2(t)=3$	$d_3(t)=1$
	$i_1(t)=i_i(t-1)$	$i_2(t)=i_2(t-1)$	$i_3(t)=X_6^3$
$t=7$	$d_1(t)=1$	$d_2(t)=4$	$d_3(t)=2$
	$i_1(t)=X_7^1$	$i_2(t)=i_2(t-1)$	$i_3(t)=i_3(t-1)$
$t=8$	$d_1(t)=2$	$d_2(t)=1$	$d_3(t)=3$
	$i_1(t)=i_1(t-1)$	$i_2(t)=X_8^2$	$i_3(t)=i_3(t-1)$
$t=9$	$d_1(t)=3$	$d_2(t)=1$	$d_3(t)=4$
	$i_1(t)=i_1(t-1)$	$i_2(t)=X_9^2$	$i_3(t)=i_3(t-1)$
$t=10$	$d_1(t)=1$	$d_2(t)=2$	$d_3(t)=5$
	$i_1(t)=X_{10}^1$	$i_2(t)=i_2(t-1)$	$i_3(t)=i_3(t-1)$

$$\underbrace{\underline{d}(t)=(d_1(t),\ldots,d_K(t)),}_{\text{arm delays}} \qquad \underbrace{\underline{i}(t)=(i_1(t),\ldots,i_K(t))}_{\text{last observed states of the arms}}$$

A New Notion of State

Time	Delay of movie 1	Delay of movie 2	Delay of movie 3
	LOS of movie 1	LOS of movie 2	LOS of movie 3
$t = 3$	$d_1(t) = 3$	$d_2(t) = 2$	$d_3(t) = 1$
	$i_1(t) = X_0^1$	$i_2(t) = X_1^2$	$i_3(t) = X_2^3$
$t = 4$	$d_1(t) = 4$	$d_2(t) = 1$	$d_3(t) = 2$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_3^2$	$i_3(t) = i_3(t-1)$
$t = 5$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 3$
	$i_1(t) = X_5^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 6$	$d_1(t) = 2$	$d_2(t) = 3$	$d_3(t) = 1$
	$i_1(t) = i_i(t-1)$	$i_2(t) = i_2(t-1)$	$i_3(t) = X_6^3$
$t = 7$	$d_1(t) = 1$	$d_2(t) = 4$	$d_3(t) = 2$
	$i_1(t) = X_7^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$
$t = 8$	$d_1(t) = 2$	$d_2(t) = 1$	$d_3(t) = 3$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_8^2$	$i_3(t) = i_3(t-1)$
$t = 9$	$d_1(t) = 3$	$d_2(t) = 1$	$d_3(t) = 4$
	$i_1(t) = i_1(t-1)$	$i_2(t) = X_9^2$	$i_3(t) = i_3(t-1)$
$t = 10$	$d_1(t) = 1$	$d_2(t) = 2$	$d_3(t) = 5$
	$i_1(t) = X_{10}^1$	$i_2(t) = i_2(t-1)$	$i_3(t) = i_3(t-1)$

$$\underbrace{\underline{d}(t) = (d_1(t), \ldots, d_K(t)),}_{\text{arm delays}} \qquad \underbrace{\underline{i}(t) = (i_1(t), \ldots, i_K(t))}_{\text{last observed states of the arms}}$$

$$(B_0, A_0, X_0^{A_0}, B_1, A_1, X_1^{A_1}, \ldots, B_t, A_t, X_t^{A_t}) \equiv \{B_s, \ (\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}$$

A New Notion of State

Time	Delay of movie 1 LOS of movie 1	Delay of movie 2 LOS of movie 2	Delay of movie 3 LOS of movie 3
$t=3$	$d_1(t)=3$ $i_1(t)=X_0^1$	$d_2(t)=2$ $i_2(t)=X_1^2$	$d_3(t)=1$ $i_3(t)=X_2^3$
$t=4$	$d_1(t)=4$ $i_1(t)=i_1(t-1)$	$d_2(t)=1$ $i_2(t)=X_3^2$	$d_3(t)=2$ $i_3(t)=i_3(t-1)$
$t=5$	$d_1(t)=1$ $i_1(t)=X_5^1$	$d_2(t)=2$ $i_2(t)=i_2(t-1)$	$d_3(t)=3$ $i_3(t)=i_3(t-1)$
$t=6$	$d_1(t)=2$ $i_1(t)=i_i(t-1)$	$d_2(t)=3$ $i_2(t)=i_2(t-1)$	$d_3(t)=1$ $i_3(t)=X_6^3$
$t=7$	$d_1(t)=1$ $i_1(t)=X_7^1$	$d_2(t)=4$ $i_2(t)=i_2(t-1)$	$d_3(t)=2$ $i_3(t)=i_3(t-1)$
$t=8$	$d_1(t)=2$ $i_1(t)=i_1(t-1)$	$d_2(t)=1$ $i_2(t)=X_8^2$	$d_3(t)=3$ $i_3(t)=i_3(t-1)$
$t=9$	$d_1(t)=3$ $i_1(t)=i_1(t-1)$	$d_2(t)=1$ $i_2(t)=X_9^2$	$d_3(t)=4$ $i_3(t)=i_3(t-1)$
$t=10$	$d_1(t)=1$ $i_1(t)=X_{10}^1$	$d_2(t)=2$ $i_2(t)=i_2(t-1)$	$d_3(t)=5$ $i_3(t)=i_3(t-1)$

$$\underbrace{\underline{d}(t)=(d_1(t),\ldots,d_K(t))}_{\text{arm delays}}, \qquad \underbrace{\underline{i}(t)=(i_1(t),\ldots,i_K(t))}_{\text{last observed states of the arms}}$$

$$(B_0, A_0, X_0^{A_0}, B_1, A_1, X_1^{A_1}, \ldots, B_t, A_t, X_t^{A_t}) \equiv \{B_s,\ (\underline{d}(s), \underline{i}(s)) : K \le s \le t\}$$

A interplay of the various variables:

$$\{B_s,\ (\underline{d}(s), \underline{i}(s)) : K \le s \le t-1\} \longrightarrow B_t \xrightarrow{\textbf{TH}} (A_t,\ X_t^{A_t}) \longrightarrow (\underline{d}(t), \underline{i}(t))$$

A Controlled Markov Process

For all t,

$$P_h(\underline{d}(t+1), \underline{i}(t+1) \mid B_t,\ \{B_s,\ (\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\})$$
$$= P_h(\underline{d}(t+1), \underline{i}(t+1) \mid B_t,\ (\underline{d}(t), \underline{i}(t)))$$

A Controlled Markov Process

For all t,

$$P_h(\underline{d}(t+1), \underline{i}(t+1) \mid B_t,\ \{B_s,\ (\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\})$$
$$= P_h(\underline{d}(t+1), \underline{i}(t+1) \mid B_t,\ (\underline{d}(t), \underline{i}(t)))$$

We have a Markov decision problem with

State space	Set of all possible $(\underline{d}, \underline{i})$ values
Action space	Set of arms
State at time t	$(\underline{d}(t), \underline{i}(t))$
Action at time t	B_t
Observation at time t	$(A_t,\ X_t^{A_t})$

A Controlled Markov Process

For all t,

$$P_h(\underline{d}(t+1), \underline{i}(t+1) \mid B_t,\ \{B_s,\ (\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\})$$
$$= P_h(\underline{d}(t+1), \underline{i}(t+1) \mid B_t,\ (\underline{d}(t), \underline{i}(t)))$$

We have a Markov decision problem with

State space	Set of all possible $(\underline{d}, \underline{i})$ values
Action space	Set of arms
State at time t	$(\underline{d}(t), \underline{i}(t))$
Action at time t	B_t
Observation at time t	$(A_t,\ X_t^{A_t})$

Our objective: characterise

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{\mathbb{E}_h[\tau(\pi)]}{\log(1/\epsilon)}$$

SRS Policies

Recall: a policy π is a mapping

$$\{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\} \mapsto B_t$$

[1] V. S. Borkar, "Control of markov chains with long-run average cost criterion," in Stochastic Differential Systems, Stochastic Control Theory and Applications. Springer, 1988, pp. 57–77.

SRS Policies

Recall: a policy π is a mapping

$$\{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\} \mapsto B_t$$

Following terminology from Borkar[1], we say π is a **stationary randomised strategy (SRS)** if there exists $\lambda(\cdot \mid \cdot)$ such that for all t,

$$P(B_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \lambda(a \mid \underline{d}(t), \underline{i}(t)).$$

[1] V. S. Borkar, "Control of markov chains with long-run average cost criterion," in Stochastic Differential Systems, Stochastic Control Theory and Applications. Springer, 1988, pp. 57–77.

SRS Policies

Recall: a policy π is a mapping

$$\{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\} \mapsto B_t$$

Following terminology from Borkar[1], we say π is a **stationary randomised strategy (SRS)** if there exists $\lambda(\cdot \mid \cdot)$ such that for all t,

$$P(B_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \lambda(a \mid \underline{d}(t), \underline{i}(t)).$$

Note that

$$P(A_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \frac{\eta}{K} + (1 - \eta)\, \lambda(a \mid \underline{d}(t), \underline{i}(t))$$

[1] V. S. Borkar, "Control of markov chains with long-run average cost criterion," in Stochastic Differential Systems, Stochastic Control Theory and Applications. Springer, 1988, pp. 57–77.

SRS Policies

Recall: a policy π is a mapping

$$\{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\} \mapsto B_t$$

Following terminology from Borkar[1], we say π is a **stationary randomised strategy (SRS)** if there exists $\lambda(\cdot \mid \cdot)$ such that for all t,

$$P(B_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \lambda(a \mid \underline{d}(t), \underline{i}(t)).$$

Note that

$$P(A_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \frac{\eta}{K} + (1 - \eta)\, \lambda(a \mid \underline{d}(t), \underline{i}(t))$$

Call such a SRS policy π^λ

[1] V. S. Borkar, "Control of markov chains with long-run average cost criterion," in Stochastic Differential Systems, Stochastic Control Theory and Applications. Springer, 1988, pp. 57–77.

SRS Policies

Recall: a policy π is a mapping

$$\{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\} \mapsto B_t$$

Following terminology from Borkar[1], we say π is a **stationary randomised strategy (SRS)** if there exists $\lambda(\cdot \mid \cdot)$ such that for all t,

$$P(B_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \lambda(a \mid \underline{d}(t), \underline{i}(t)).$$

Note that

$$P(A_t = a \mid \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \frac{\eta}{K} + (1 - \eta)\, \lambda(a \mid \underline{d}(t), \underline{i}(t))$$

Call such a SRS policy π^λ

Π_{SRS} = set of all SRS policies

[1] V. S. Borkar, "Control of markov chains with long-run average cost criterion," in Stochastic Differential Systems, Stochastic Control Theory and Applications. Springer, 1988, pp. 57–77.

SRS Policies + Trembling Hand = Ergodicity

A Key Ergodicity Property

Under any $\pi^\lambda \in \Pi_{\mathsf{SRS}}$, the CMP $\{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}$ is a Markov process. Further, if the trembling hand parameter $\eta > 0$, then this Markov process is ergodic.

A unique stationary distribution, call it μ^λ, therefore exists under π^λ

SRS Policies + Trembling Hand = Ergodicity

A Key Ergodicity Property

Under any $\pi^\lambda \in \Pi_{\mathsf{SRS}}$, the CMP $\{(\underline{d}(s), \underline{i}(s)) : K \le s \le t\}$ is a Markov process. Further, if the trembling hand parameter $\eta > 0$, then this Markov process is ergodic.

A unique stationary distribution, call it μ^λ, therefore exists under π^λ

Ergodic state action occupancy measure:

$$\nu^\lambda(\underline{d}, \underline{i}, a) = \mu^\lambda(\underline{d}, \underline{i}) \ \left(\frac{\eta}{K} + (1 - \eta)\, \lambda(a \mid \underline{d}, \underline{i}) \right)$$

Lower Bound - 1

Fix odd arm h ($\mathcal{H}_h$ is the true hypothesis)
P_1: TPM of arm h
P_2: TPM of arm $h' \neq h$

Lower Bound - 1

Fix odd arm h ($\mathcal{H}_h$ is the true hypothesis)
P_1: TPM of arm h
P_2: TPM of arm $h' \neq h$

Let

$$P_h^a \;=\; \text{TPM of arm } a \text{ under hypothesis } \mathcal{H}_h$$

For any integer $d \geq 1$,

$$(P_h^a)^d \;=\; d\text{th power of TPM } P_h^a$$

Lower Bound - 2

Lower Bound: Restless Markov Arms

$$\liminf_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E_h[\tau(\pi)]}{\log(1/\epsilon)} \geq \frac{1}{R^*(P_1, P_2)},$$

where

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{\mathrm{SRS}}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i})} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a) \, D((P_h^a)^{d_a}(\cdot|i_a) \| (P_{h'}^a)^{d_a}(\cdot|i_a))$$

Lower Bound - 2

Lower Bound: Restless Markov Arms

$$\liminf_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E_h[\tau(\pi)]}{\log(1/\epsilon)} \geq \frac{1}{R^*(P_1, P_2)},$$

where

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{\mathsf{SRS}}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i})} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\, D((P_h^a)^{d_a}(\cdot|i_a) \| (P_{h'}^a)^{d_a}(\cdot|i_a))$$

Remarks:

- $R^*(P_1, P_2)$ does not depend on h, the location of odd arm
- $R^*(P_1, P_2)$: metric for closeness between odd and the non-odd movies
- Computability of $R^*(P_1, P_2)$: Q-learning or other such algorithms

Achievability

Preliminaries

- $R^*(P_1, P_2)$ has a sup in it
- It is not clear if this sup is achievable by some SRS policy

Preliminaries

- $R^*(P_1, P_2)$ has a sup in it
- It is not clear if this sup is achievable by some SRS policy
- However, the sup may be approached arbitrarily closely:

$$\forall\ \delta > 0,\ \exists\ \lambda_{h,\delta}(\cdot \mid \cdot) \text{ s.t.}$$

$$\min_{h' \neq h} \sum_{(\underline{d},\underline{i})} \sum_{a=1}^{K} \nu^{\lambda_{h,\delta}}(\underline{d}, \underline{i}, a)\, D((P_h^a)^{d_a}(\cdot|i_a) \| (P_{h'}^a)^{d_a}(\cdot|i_a)) > \frac{R^*(P_1, P_2)}{1+\delta}$$

Achievability: Policy

Policy $\pi^{\star}(L, \delta)$

Achievability

- Input: Two parameters $L > 1$ and $\delta > 0$

Achievability: Policy

Policy $\pi^\star(L, \delta)$

Achievability

- Input: Two parameters $L > 1$ and $\delta > 0$
- Select arm 1 at time $t = 0$, arm 2 at time $t = 1$ and so on until arm K at time $t = K - 1$

Achievability: Policy

Policy $\pi^\star(L, \delta)$

Achievability

- Input: Two parameters $L > 1$ and $\delta > 0$
- Select arm 1 at time $t = 0$, arm 2 at time $t = 1$ and so on until arm K at time $t = K - 1$
- For all $t \geq K$:
  - Maintain guess of odd arm:

$$\hat{\theta}(n) \in \arg\max_{h} \underbrace{\min_{h' \neq h} \log \frac{P_h(B_0, A_0, X_0^{A_0}, \ldots, B_n, A_n, X_n^{A_n})}{P_{h'}(B_0, A_0, X_0^{A_0}, \ldots, B_n, A_n, X_n^{A_n})}}_{M_h(n)}$$

Achievability: Policy

Policy $\pi^\star(L, \delta)$

Achievability

- Input: Two parameters $L > 1$ and $\delta > 0$
- Select arm 1 at time $t = 0$, arm 2 at time $t = 1$ and so on until arm K at time $t = K - 1$
- For all $t \geq K$:
  - Maintain guess of odd arm:

$$\hat{\theta}(n) \in \arg\max_{h} \underbrace{\min_{h' \neq h} \log \frac{P_h(B_0, A_0, X_0^{A_0}, \ldots, B_n, A_n, X_n^{A_n})}{P_{h'}(B_0, A_0, X_0^{A_0}, \ldots, B_n, A_n, X_n^{A_n})}}_{M_h(n)}$$

  - If $M_{\hat{\theta}(n)}(n) \geq \log((K-1)L)$, stop and declare $\hat{\theta}(n)$ is the odd arm

Achievability: Policy

Policy $\pi^\star(L, \delta)$

Achievability

- Input: Two parameters $L > 1$ and $\delta > 0$
- Select arm 1 at time $t = 0$, arm 2 at time $t = 1$ and so on until arm K at time $t = K - 1$
- For all $t \geq K$:
  - Maintain guess of odd arm:

$$\hat{\theta}(n) \in \arg\max_h \underbrace{\min_{h' \neq h} \log \frac{P_h(B_0, A_0, X_0^{A_0}, \ldots, B_n, A_n, X_n^{A_n})}{P_{h'}(B_0, A_0, X_0^{A_0}, \ldots, B_n, A_n, X_n^{A_n})}}_{M_h(n)}$$

  - If $M_{\hat{\theta}(n)}(n) \geq \log((K-1)L)$, stop and declare $\hat{\theta}(n)$ is the odd arm
  - If $M_{\hat{\theta}(n)}(n) < \log((K-1)L)$, select next arm according to the distribution $\lambda_{\hat{\theta}(n),\delta}(\cdot \mid \cdot)$

Achievability: Results

- Policy $\pi^\star(L, \delta)$ stops in finite time w.p. 1
- If $L = 1/\epsilon$, then $\pi^\star(L, \delta) \in \Pi(\epsilon)$ for all $\delta > 0$ (desired error probability)
- **Upper bound: for $\pi = \pi^\star(L, \delta)$,**

$$\limsup_{L \to \infty} \frac{E_h[\tau(\pi)]}{\log L} \leq \frac{1 + \delta}{R^*(P_1, P_2)}$$

Achievability: Results

- Policy $\pi^\star(L, \delta)$ stops in finite time w.p. 1
- If $L = 1/\epsilon$, then $\pi^\star(L, \delta) \in \Pi(\epsilon)$ for all $\delta > 0$ (desired error probability)
- **Upper bound: for $\pi = \pi^\star(L, \delta)$,**

$$\limsup_{L \to \infty} \frac{E_h[\tau(\pi)]}{\log L} \leq \frac{1 + \delta}{R^*(P_1, P_2)}$$

- Stitch together the solutions for various $\delta > 0$:

$$\limsup_{\delta \downarrow 0} \limsup_{L \to \infty} \frac{E_h[\tau(\pi)]}{\log L} \leq \frac{1}{R^*(P_1, P_2)}$$

Main Result

Main Result

Main Result

For the problem of odd arm identification with restless Markov arms in which h is the odd arm, P_1 is the TPM of arm h and P_2 is the common TPM of all arms other than h, where $P_2 \neq P_1$,

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{\mathbb{E}_h[\tau(\pi)]}{\log \frac{1}{\epsilon}} = \frac{1}{R^*(P_1, P_2)}.$$

The analysis crucially relies on $\eta > 0$.

The Case $\eta = 0$

A Key Monotonicity Property

- Let $\Pi_{\mathsf{SRS}}^{\eta}$ be the set of all SRS policies with η-trembling hand, i.e.,

$$\Pi_{\mathsf{SRS}}^{\eta} = \left\{ \frac{\eta}{K} + (1-\eta)\ \lambda(\cdot \mid \cdot) \right\}_{\lambda}$$

A Key Monotonicity Property

- Let $\Pi_{\mathsf{SRS}}^{\eta}$ be the set of all SRS policies with η-trembling hand, i.e.,

$$\Pi_{\mathsf{SRS}}^{\eta} = \left\{ \frac{\eta}{K} + (1-\eta)\ \lambda(\cdot \mid \cdot) \right\}_{\lambda}$$

A Key Monotonicity Property

If $0 < \eta' < \eta \leq 1$, then $\Pi_{\mathsf{SRS}}^{\eta} \subseteq \Pi_{\mathsf{SRS}}^{\eta'}$.

A Key Monotonicity Property

- Let $\Pi^{\eta}_{\mathsf{SRS}}$ be the set of all SRS policies with η-trembling hand, i.e.,

$$\Pi^{\eta}_{\mathsf{SRS}} = \left\{ \frac{\eta}{K} + (1-\eta)\ \lambda(\cdot \mid \cdot) \right\}_{\lambda}$$

A Key Monotonicity Property

If $0 < \eta' < \eta \leq 1$, then $\Pi^{\eta}_{\mathsf{SRS}} \subseteq \Pi^{\eta'}_{\mathsf{SRS}}$.

- Relabel $R^*(P_1, P_2)$ from before as $R^*_{\eta}(P_1, P_2)$

A Key Monotonicity Property

- Let $\Pi^{\eta}_{\mathsf{SRS}}$ be the set of all SRS policies with η-trembling hand, i.e.,

$$\Pi^{\eta}_{\mathsf{SRS}} = \left\{ \frac{\eta}{K} + (1-\eta)\ \lambda(\cdot \mid \cdot) \right\}_{\lambda}$$

A Key Monotonicity Property

If $0 < \eta' < \eta \leq 1$, then $\Pi^{\eta}_{\mathsf{SRS}} \subseteq \Pi^{\eta'}_{\mathsf{SRS}}$.

- Relabel $R^*(P_1, P_2)$ from before as $R^*_{\eta}(P_1, P_2)$
- From the above monotonicity, it follows that $R^*_{\eta}(P_1, P_2)$ **is non-decreasing in** η

Gap or No Gap

- Define

$$\Pi^0_{\mathsf{SRS}} = \Pi^{\eta=0}_{\mathsf{SRS}}, \qquad R^*_0(P_1, P_2) = R^*_{\eta=0}(P_1, P_2)$$

- We then have

$$\lim_{\eta \downarrow 0} R^*_\eta(P_1, P_2) \le R^*_0(P_1, P_2)$$

Is the above inequality an equality? (right continuity at $\eta = 0$)

Gap or No Gap

- Define

$$\Pi_{\mathsf{SRS}}^{0} = \Pi_{\mathsf{SRS}}^{\eta=0}, \qquad R_0^*(P_1, P_2) = R_{\eta=0}^*(P_1, P_2)$$

- We then have

$$\lim_{\eta \downarrow 0} R_\eta^*(P_1, P_2) \le R_0^*(P_1, P_2)$$

Is the above inequality an equality? (right continuity at $\eta = 0$)

- Prior works = above inequality is an equality (Envelope theorem)

Lower Bound: IID Observations From Arms

- ν_1 = law of odd arm iid process, ν_2 = law of each of the non-odd iid processes

Lower Bound: IID Observations From Arms

- $\nu_1 =$ law of odd arm iid process, $\nu_2 =$ law of each of the non-odd iid processes
- Then, for all $i, j \in \mathcal{S}$ and $d \geq 1$,

$$(P_h^a)^d(j|i) = \nu_h^a(j)$$

Lower Bound: IID Observations From Arms

- ν_1 = law of odd arm iid process, ν_2 = law of each of the non-odd iid processes
- Then, for all $i, j \in \mathcal{S}$ and $d \geq 1$,

$$(P_h^a)^d(j|i) = \nu_h^a(j)$$

$$R_\eta^*(P_1, P_2)$$

$$= \sup_{\pi^\lambda \in \Pi_{\mathrm{SRS}}^\eta} \min_{h' \neq h} \sum_{a=1}^{K} \sum_{(\underline{d}, \underline{i})} \mu^\lambda(\underline{d}, \underline{i}) \left[\frac{\eta}{K} + (1-\eta)\,\lambda(a|\underline{d}, \underline{i})\right] D(\nu_h^a \| \nu_{h'}^a)$$

$$= \sup_{\lambda \in \mathcal{P}(\mathcal{A})} \min_{h' \neq h} \underbrace{\frac{\eta}{K} \sum_{a=1}^{K} D(\nu_h^a \| \nu_{h'}^a) + (1-\eta) \sum_{a=1}^{K} \lambda(a)\, D(\nu_h^a \| \nu_{h'}^a)}_{\text{absolutely continuous for all } \eta \in [0,1]}$$

Lower Bound: IID Observations From Arms

- ν_1 = law of odd arm iid process, ν_2 = law of each of the non-odd iid processes
- Then, for all $i, j \in \mathcal{S}$ and $d \geq 1$,

$$(P_h^a)^d(j|i) = \nu_h^a(j)$$

$R_\eta^*(P_1, P_2)$

$$= \sup_{\pi^\lambda \in \Pi_{\mathrm{SRS}}^\eta} \min_{h' \neq h} \sum_{a=1}^{K} \sum_{(\underline{d}, \underline{i})} \mu^\lambda(\underline{d}, \underline{i}) \left[\frac{\eta}{K} + (1-\eta)\,\lambda(a|\underline{d}, \underline{i})\right] D(\nu_h^a \| \nu_{h'}^a)$$

$$= \sup_{\lambda \in \mathcal{P}(\mathcal{A})} \min_{h' \neq h} \underbrace{\frac{\eta}{K} \sum_{a=1}^{K} D(\nu_h^a \| \nu_{h'}^a) + (1-\eta) \sum_{a=1}^{K} \lambda(a)\, D(\nu_h^a \| \nu_{h'}^a)}_{\text{absolutely continuous for all } \eta \in [0,1]}$$

- Using Envelope theorem, $R_\eta^*(P_1, P_2)$ is absolutely continuous for all $\eta \in [0, 1]$

Lower Bound: Rested Markov Arms

- Delay of all arms $\equiv 1$ for all t

Lower Bound: Rested Markov Arms

- Delay of all arms $\equiv 1$ for all t

$$
\begin{aligned}
&R^*_\eta(P_1, P_2) \\
&= \sup_{\pi^\lambda \in \Pi^\eta_{\mathsf{SRS}}} \min_{h' \neq h} \sum_{a=1}^{K} \sum_{\underline{i}} \mu^{\lambda^\eta}(\underline{i}) \left[\frac{\eta}{K} + (1-\eta)\,\lambda(a|\underline{i})\right] D(P^a_h(\cdot|i_a) \| P^a_{h'}(\cdot|i_a)) \\
&= \sup_{\pi^\lambda \in \Pi^\eta_{\mathsf{SRS}}} \min_{h' \neq h} \Bigg\{ \frac{\eta}{K} \sum_{a=1}^{K} \sum_{\underline{i}} \mu^\lambda(\underline{i})\ D(P^a_h(\cdot|i_a) \| P^a_{h'}(\cdot|i_a)) \\
&\qquad + (1-\eta) \sum_{a=1}^{K} \sum_{\underline{i}} \mu^\lambda(\underline{i})\ \lambda(a \mid \underline{i})\ D(P^a_h(\cdot|i_a) \| P^a_{h'}(\cdot|i_a)) \Bigg\} \\
&= \sup_{\lambda \in \mathcal{P}(\mathcal{A})} \min_{h' \neq h} \Bigg\{ {\color{red}\frac{\eta}{K}} \sum_{a=1}^{K} D(P^a_h(\cdot|\cdot) \| P^a_{h'}(\cdot|\cdot) | \mu^a_h) \\
&\qquad + {\color{red}(1-\eta)} \sum_{a=1}^{K} \lambda(a)\, D(P^a_h(\cdot|\cdot) \| P^a_{h'}(\cdot|\cdot) | \mu^a_h) \Bigg\}
\end{aligned}
$$

Concluding Remarks

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms
- A sequence of policies that come arbitrarily close to the lower bound asymptotically

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms
- A sequence of policies that come arbitrarily close to the lower bound asymptotically
- Lower bound generalises the results of prior works

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms
- A sequence of policies that come arbitrarily close to the lower bound asymptotically
- Lower bound generalises the results of prior works
- Trembling hand crucial for analysis

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms
- A sequence of policies that come arbitrarily close to the lower bound asymptotically
- Lower bound generalises the results of prior works
- Trembling hand crucial for analysis

Some future directions:

- Without trembling hand, does Envelope theorem hold?

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms
- A sequence of policies that come arbitrarily close to the lower bound asymptotically
- Lower bound generalises the results of prior works
- Trembling hand crucial for analysis

Some future directions:

- Without trembling hand, does Envelope theorem hold?
- The case when P_1 and P_2 are not known

Concluding Remarks

- A first known asymptotic lower bound for the setting of restless Markov arms
- A sequence of policies that come arbitrarily close to the lower bound asymptotically
- Lower bound generalises the results of prior works
- Trembling hand crucial for analysis

Some future directions:

- Without trembling hand, does Envelope theorem hold?
- The case when P_1 and P_2 are not known

Thank You!