Learning to Detect an Odd Markov Arm

Karthik PN

Indian Statistical Institute, Delhi

Based on joint work with Rajesh Sundaresan

1 P_2

2 P_1

3 P_2

4 P_2

Time

Time

Obs

1 P_2

X_0^1

2 P_1

3 P_2

4 P_2

Time

Time	0													
Obs	X_0^1													

1 P_2

X_0^1

2 P_1

X_0^2

3 P_2

4 P_2

Time

Time	0	1												
Obs	X_0^1	X_0^2												

1 P_2

X_0^1

2 P_1

X_0^2

3 P_2

X_0^3

4 P_2

Time

Time	0	1	2										
Obs	X_0^1	X_0^2	X_0^3										

1 P_2

X_0^1

2 P_1

X_0^2

3 P_2

X_0^3

4 P_2

X_0^4

Time

Time	0	1	2	3
Obs	X_0^1	X_0^2	X_0^3	X_0^4

1 P_2

X_0^1

2 P_1

X_0^2 X_1^2

3 P_2

X_0^3

4 P_2

X_0^4

Time

Time	0	1	2	3	4
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2

1 P_2

X_0^1

2 P_1

X_0^2 X_1^2

3 P_2

X_0^3 X_1^3

4 P_2

X_0^4

Time

Time	0	1	2	3	4	5								
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3								

1 P_2

X_0^1

2 P_1

X_0^2 X_1^2 X_2^2

3 P_2

X_0^3 X_1^3

4 P_2

X_0^4

Time

Time	0	1	2	3	4	5	6
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2

1 P_2

X_0^1 X_1^1

2 P_1

X_0^2 X_1^2 X_2^2

3 P_2

X_0^3 X_1^3

4 P_2

X_0^4

Time

Time	0	1	2	3	4	5	6	7							
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1							

1 P_2

X_0^1 X_1^1

2 P_1

X_0^2 X_1^2 X_2^2

3 P_2

X_0^3 X_1^3

4 P_2

X_0^4 X_1^4

Time

Time	0	1	2	3	4	5	6	7	8
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1	X_1^4

1 P_2

X_0^1 X_1^1

2 P_1

X_0^2 X_1^2 X_2^2 X_3^2

3 P_2

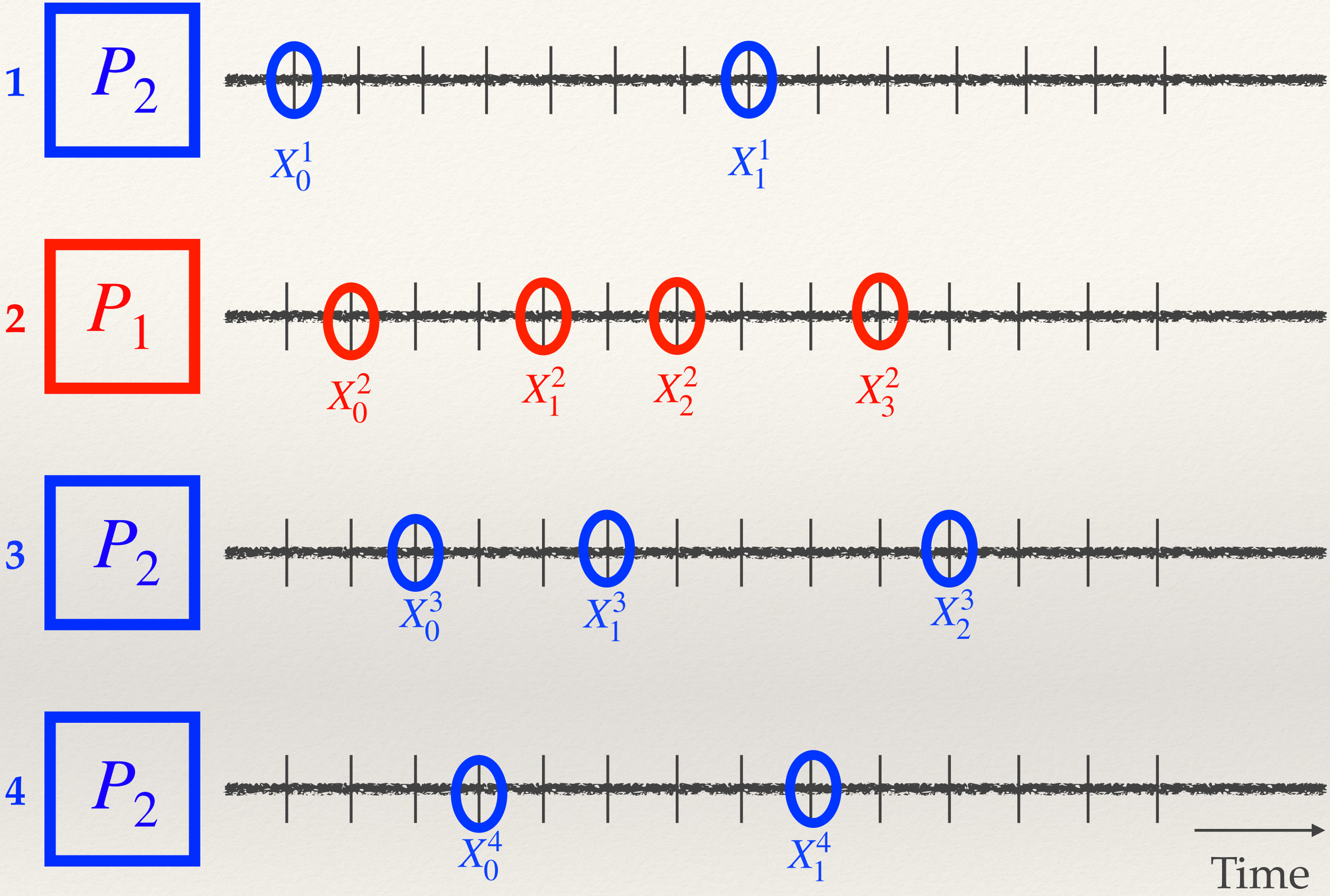
X_0^3 X_1^3

4 P_2

X_0^4 X_1^4

Time

Time	0	1	2	3	4	5	6	7	8	9
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1	X_1^4	X_3^2



Time	0	1	2	3	4	5	6	7	8	9	10				
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1	X_1^4	X_3^2	X_2^3				

1 P_2

X_0^1 X_1^1

2 P_1

X_0^2 X_1^2 X_2^2 X_3^2 X_4^2

3 P_2

X_0^3 X_1^3 X_2^3

4 P_2

X_0^4 X_1^4

Time

Time	0	1	2	3	4	5	6	7	8	9	10	11			
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1	X_1^4	X_3^2	X_2^3	X_4^2			

1 P_2

X_0^1 X_1^1

2 P_1

X_0^2 X_1^2 X_2^2 X_3^2 X_4^2 X_5^2

3 P_2

X_0^3 X_1^3 X_2^3

4 P_2

X_0^4 X_1^4

Time

Time	0	1	2	3	4	5	6	7	8	9	10	11	12		
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1	X_1^4	X_3^2	X_2^3	X_4^2	X_5^2		

1 P_2

X_0^1 X_1^1

2 P_1

X_0^2 X_1^2 X_2^2 X_3^2 X_4^2 X_5^2

3 P_2

X_0^3 X_1^3 X_2^3 X_3^3

4 P_2

X_0^4 X_1^4

Time

Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	
Obs	X_0^1	X_0^2	X_0^3	X_0^4	X_1^2	X_1^3	X_2^2	X_1^1	X_1^4	X_3^2	X_2^3	X_4^2	X_5^2	X_3^3	

- Based on history

$$\mathscr{F}_t = (A_0, \bar{X}_0, A_1, \bar{X}_1, \ldots, A_t, \bar{X}_t)$$

the learner should pick A_{t+1} (continue)

OR

the learner should stop and declare
the index of the "odd" arm (arm with TPM P_1) (stop)

- Learner knows neither P_1 nor P_2

- Based on history

$$\mathscr{F}_t = (A_0, \bar{X}_0, A_1, \bar{X}_1, \ldots, A_t, \bar{X}_t)$$

the learner should pick A_{t+1} (continue)

OR

the learner should stop and declare
the index of the "odd" arm (arm with TPM P_1) (stop)

- Learner knows neither P_1 nor P_2

Policy of the learner

- Based on history

$$\mathscr{F}_t = (A_0, \bar{X}_0, A_1, \bar{X}_1, \ldots, A_t, \bar{X}_t)$$

the learner should pick A_{t+1} (continue)

OR

the learner should stop and declare
the index of the "odd" arm (arm with TPM P_1) (stop)

- Learner knows neither P_1 nor P_2

Policy of the learner

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

- $C = (h, P_1, P_2)$: a configuration of the arms with

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,
  - P_1: TPM of odd arm

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,
  - P_1: TPM of odd arm
  - P_2: TPM of each of the non-odd arms

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,
  - P_1: TPM of odd arm
  - P_2: TPM of each of the non-odd arms
- Two quantities of interest:

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,
  - P_1: TPM of odd arm
  - P_2: TPM of each of the non-odd arms
- Two quantities of interest:
  - $E^{\pi}[\tau(\pi) \mid C]$

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,
  - P_1: TPM of odd arm
  - P_2: TPM of each of the non-odd arms
- Two quantities of interest:
  - $E^{\pi}[\tau(\pi) \mid C]$
  - $P^{\pi}(I(\pi) \neq h \mid C) \leq \epsilon \quad \forall C = (h, P_1, P_2)$

- $C = (h, P_1, P_2)$: a configuration of the arms with
  - odd arm index h,
  - P_1: TPM of odd arm
  - P_2: TPM of each of the non-odd arms
- Two quantities of interest:
  - $E^{\pi}[\tau(\pi) \,|\, C]$
  - $P^{\pi}(I(\pi) \neq h \,|\, C) \leq \epsilon \quad \forall C = (h, P_1, P_2)$
- We will let $\epsilon \downarrow 0$ and study the behaviour of $E^{\pi}[\tau(\pi) \,|\, C]$ in this asymptotic regime

Our Results

- A first known asymptotic lower bound for the rested Markov setting

Our Results

- A first known asymptotic lower bound for the rested Markov setting

$$E^{\pi}[\tau(\pi) \,|\, C = (h, P_1, P_2)] \gtrapprox \frac{\log 1/\epsilon}{D^*(h, P_1, P_2)}$$

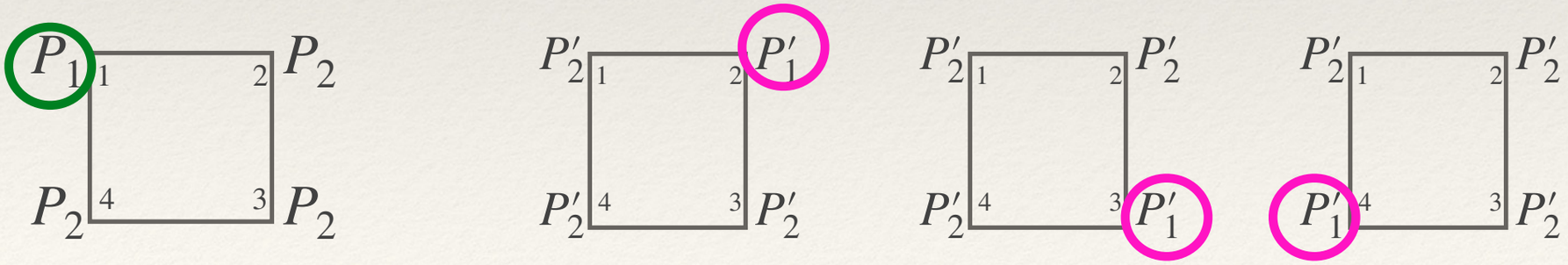
$$D^*(h, P_1, P_2) = \max_{\lambda} \min_{C'=(h', P_1', P_2'),\ h' \neq h} \sum_{a=1}^{K} \lambda(a)\, D(P_h^a \| P_{h'}^a \,|\, \mu_h^a)$$

Our Results

- A first known asymptotic lower bound for the rested Markov setting

$$E^{\pi}[\tau(\pi) \,|\, C = (h, P_1, P_2)] \gtrapprox \frac{\log 1/\epsilon}{D^*(h, P_1, P_2)}$$

$$D^*(h, P_1, P_2) = \max_{\lambda} \min_{C'=(h', P_1', P_2'),\ h' \neq h} \sum_{a=1}^{K} \lambda(a)\, D(P_h^a \| P_{h'}^a \,|\, \mu_h^a)$$



True

Possible Alternatives

Our Results

- A first known asymptotic lower bound for the rested Markov setting

$$E^{\pi}[\tau(\pi) \,|\, C = (h, P_1, P_2)] \gtrapprox \frac{\log 1/\epsilon}{D^*(h, P_1, P_2)}$$

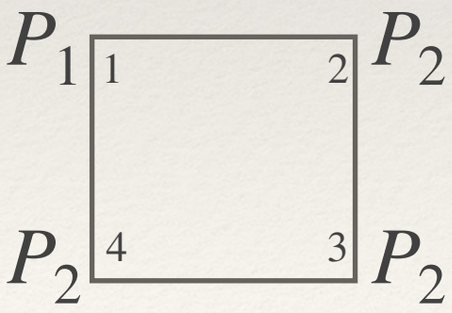
$$D^*(h, P_1, P_2) = \max_{\lambda} \min_{C'=(h', P_1', P_2'),\ h' \neq h} \sum_{a=1}^{K} \lambda(a)\, D(P_h^a \| P_{h'}^a \,|\, \mu_h^a)$$

Our Results

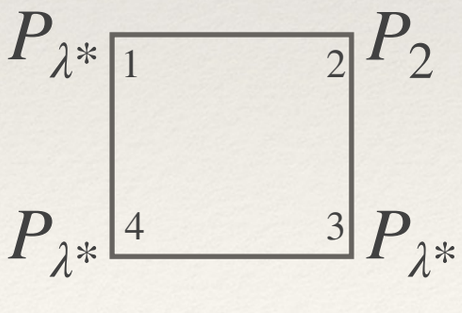
- A first known asymptotic lower bound for the rested Markov setting

$$E^{\pi}[\tau(\pi) \,|\, C = (h, P_1, P_2)] \gtrapprox \frac{\log 1/\epsilon}{D^*(h, P_1, P_2)}$$

$$D^*(h, P_1, P_2) = \max_{\lambda} \min_{C'=(h', P_1', P_2'),\ h' \neq h} \sum_{a=1}^{K} \lambda(a)\, D(P_h^a \| P_{h'}^a \,|\, \mu_h^a)$$



True



`Nearest' Alternative

P_{λ^*} = **some linear combination of P_1 and P_2**

Our Results

- A policy that meets the lower bound as $\epsilon \downarrow 0$
- Policy = modification of classical GLRT + forced sampling of arms
- Key challenges in the rested Markov setting identified