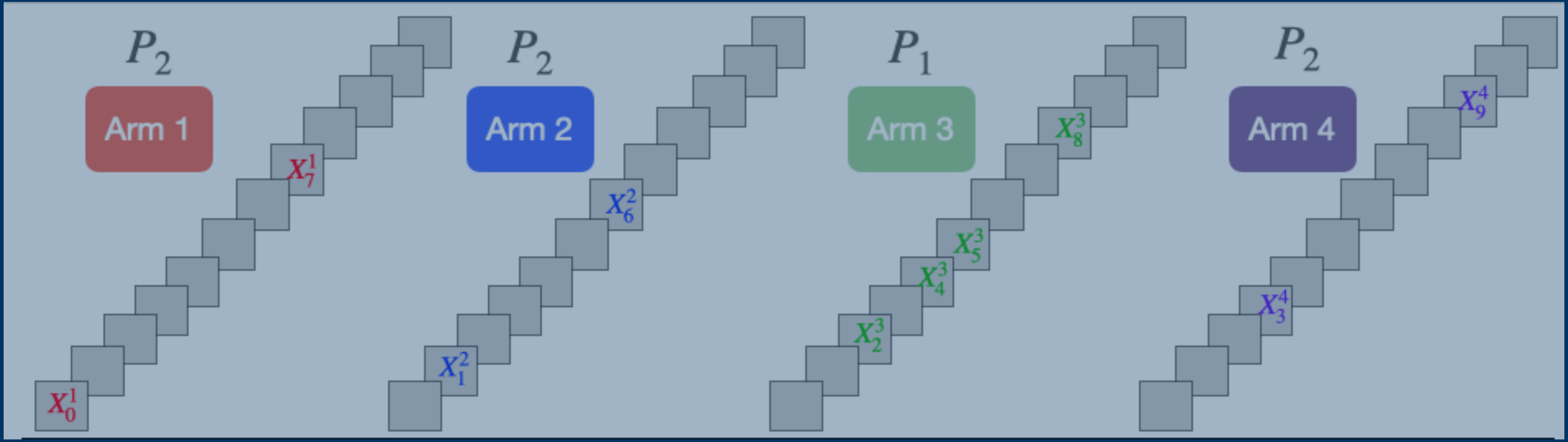


Sequential Controlled Sensing to Detect an Anomalous Process

Ph. D. Colloquium
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

P. N. Karthik

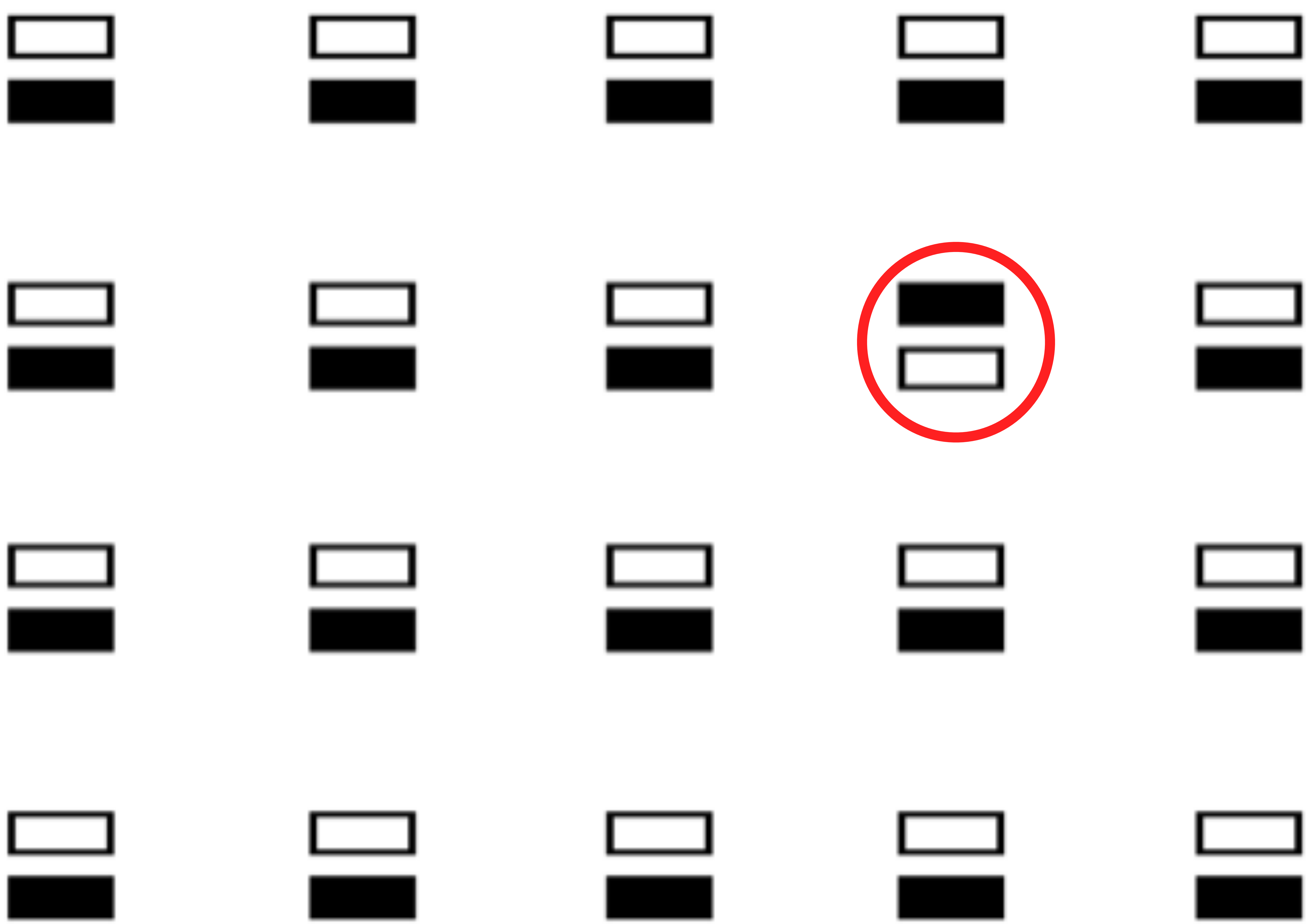
Advisor : Prof. Rajesh Sundaresan

23 June 2021

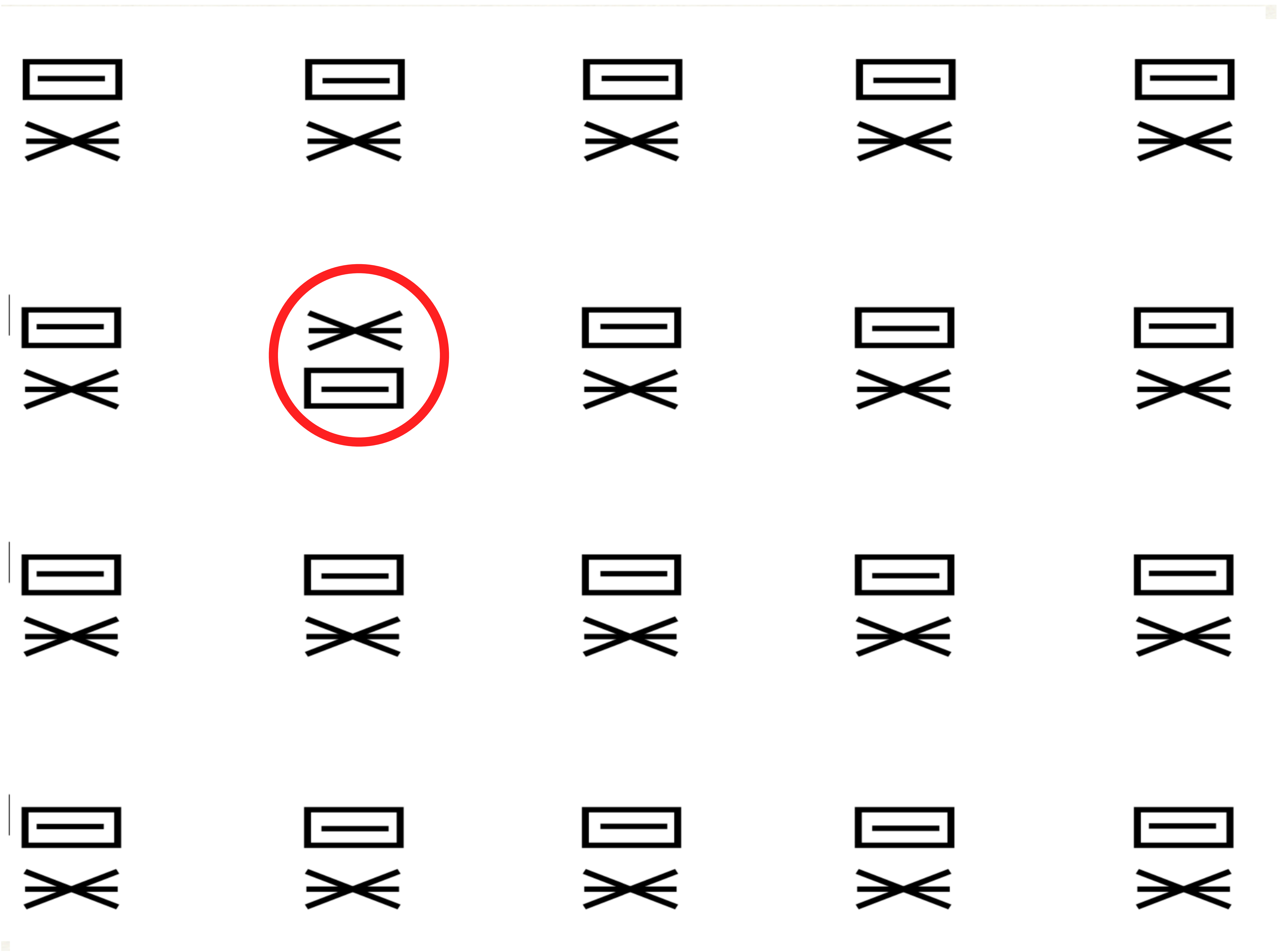
Motivation

Visual Search Experiments, Multi-Armed Bandits

Find the “Odd” Image — 1

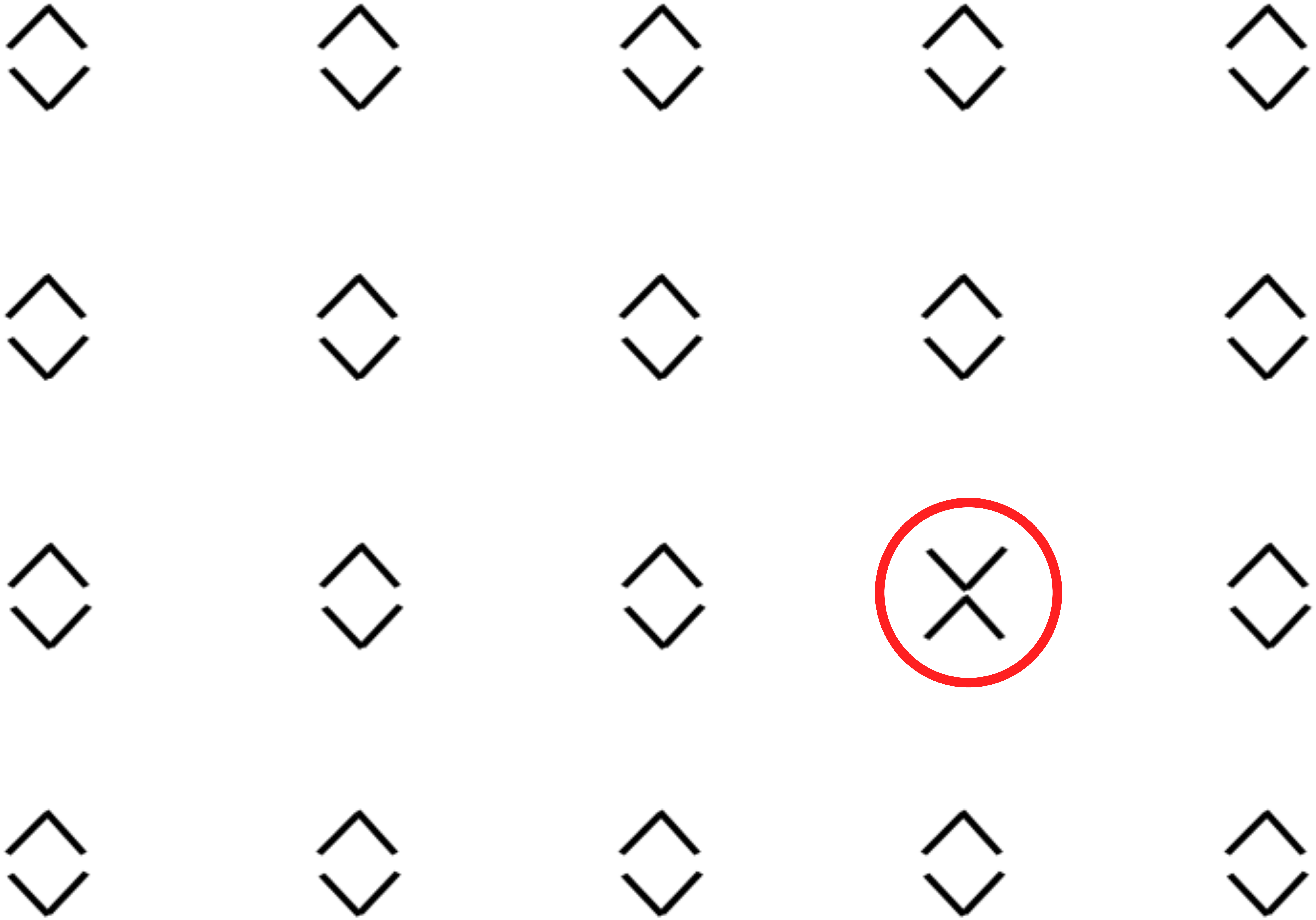


Find the “Odd” Image — 2



Odd image

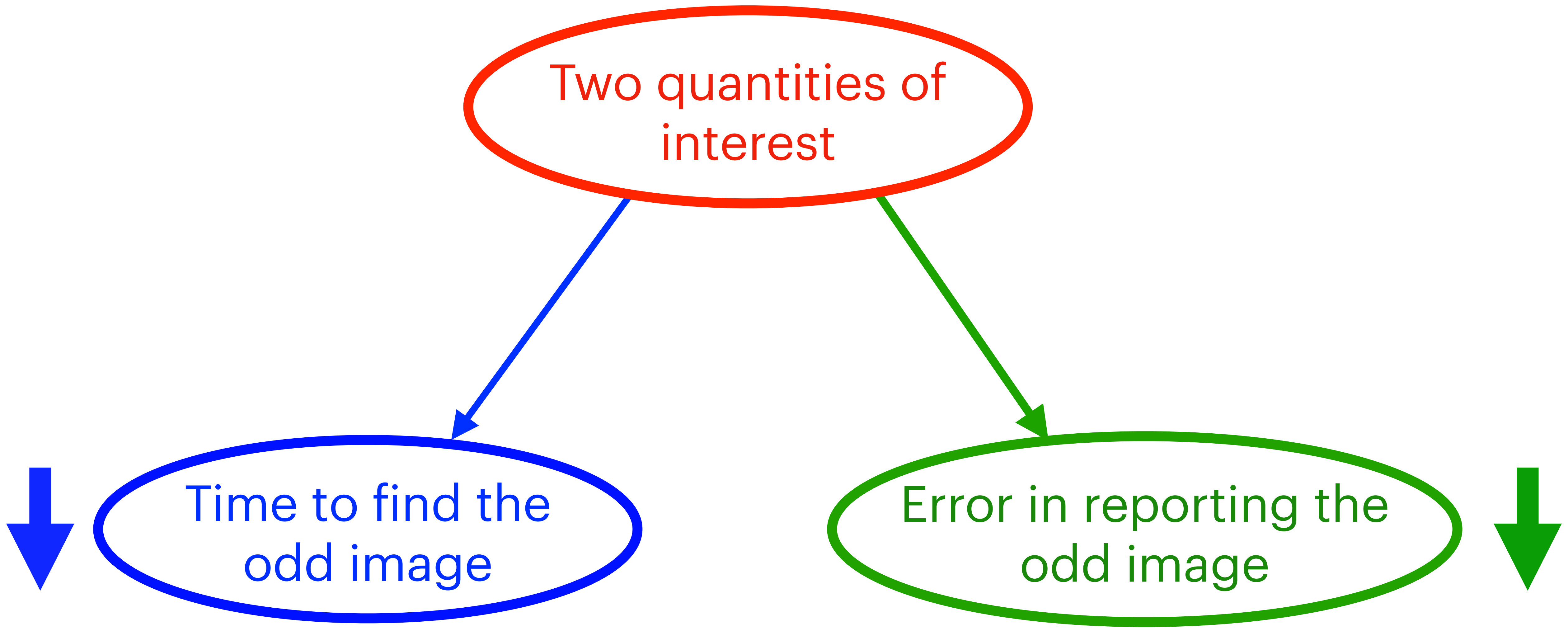
Find the “Odd” Image — 3



Odd image

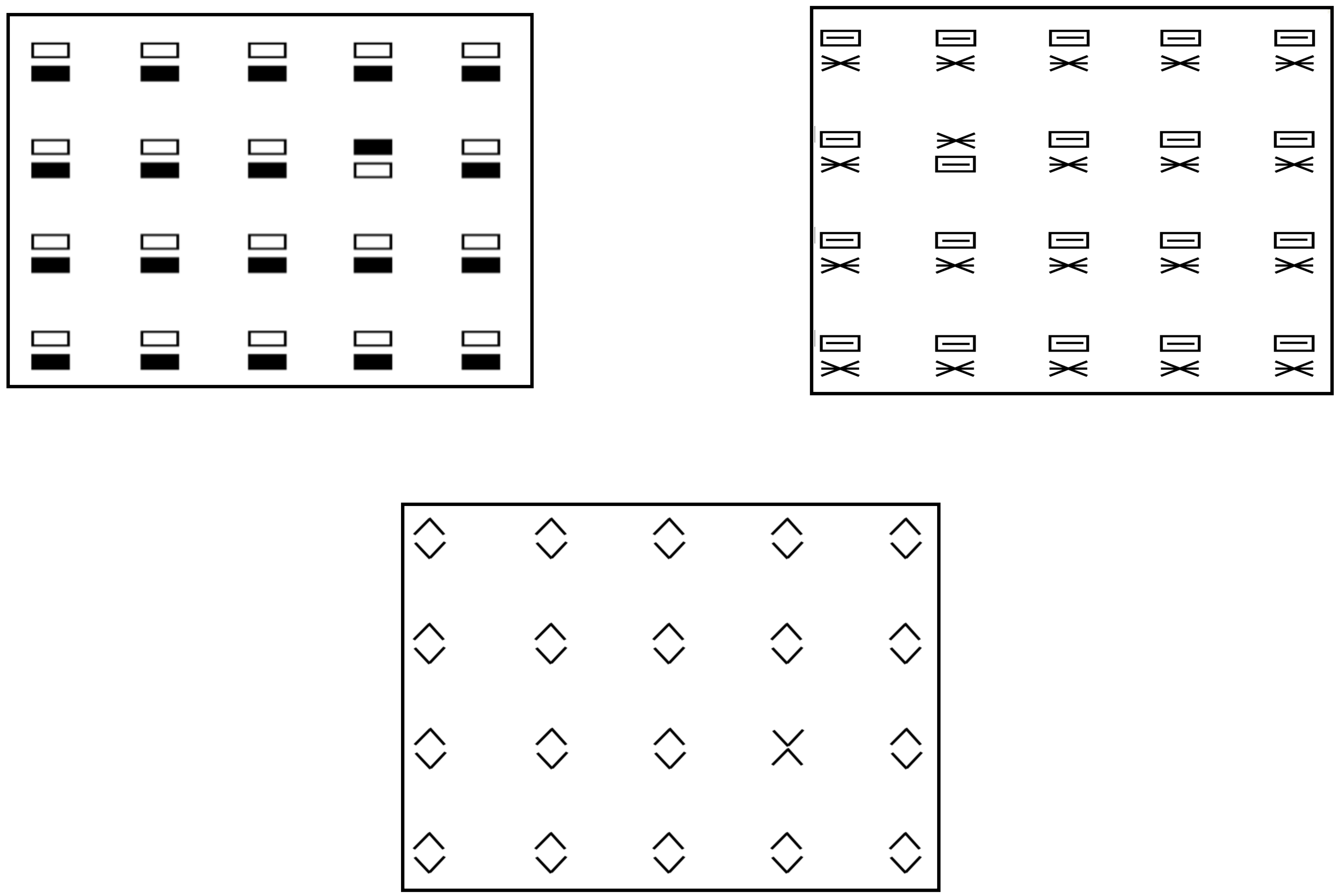
Finding the Odd Image

- Time to find the odd image depends on the image pairs
- The “closer” the image pairs are to the eyes, the longer it takes to find the odd image



Fix error and characterise the time to find odd arm as a function of error

Fixed confidence regime

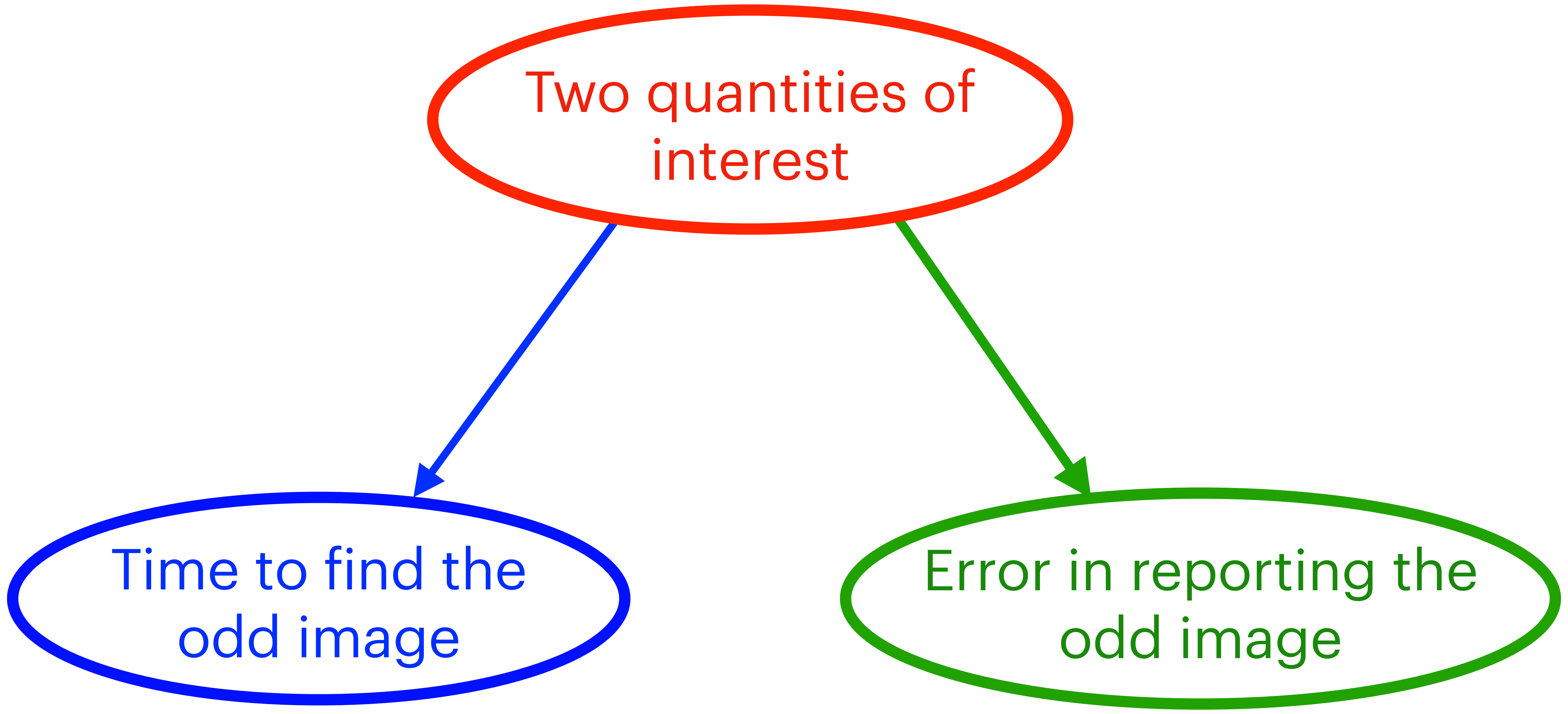


Goal: to find the odd image quickly and accurately

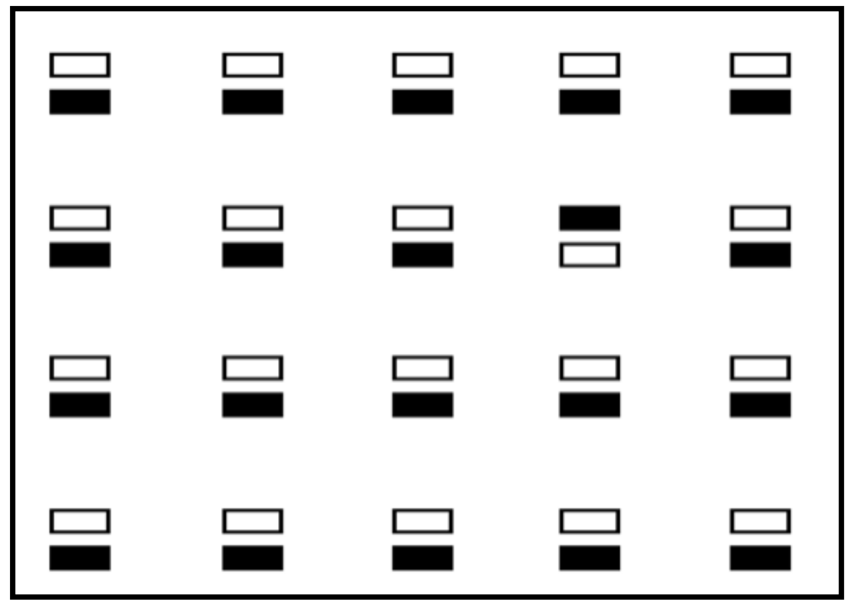
- π : strategy to find the odd image
- $\tau(\pi)$: time to find the odd image under π
- For $\epsilon \in (0,1)$, let $\Pi(\epsilon) = \{\pi : P_{\text{error}}(\pi) \leq \epsilon\}$
- Vaidhiyan et al.[1,2] showed that for any two image pairs I_1 and I_2,

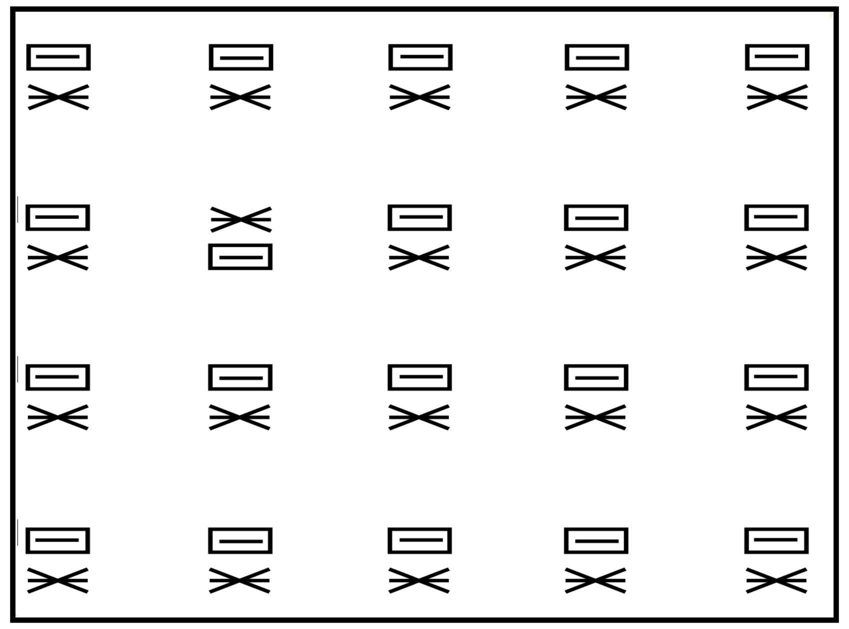
$$\inf_{\pi \in \Pi(\epsilon)} E[\tau(\pi) \,|\, I_1, I_2] \approx \alpha(I_1, I_2) \cdot \left(\log \frac{1}{\epsilon} \right)$$

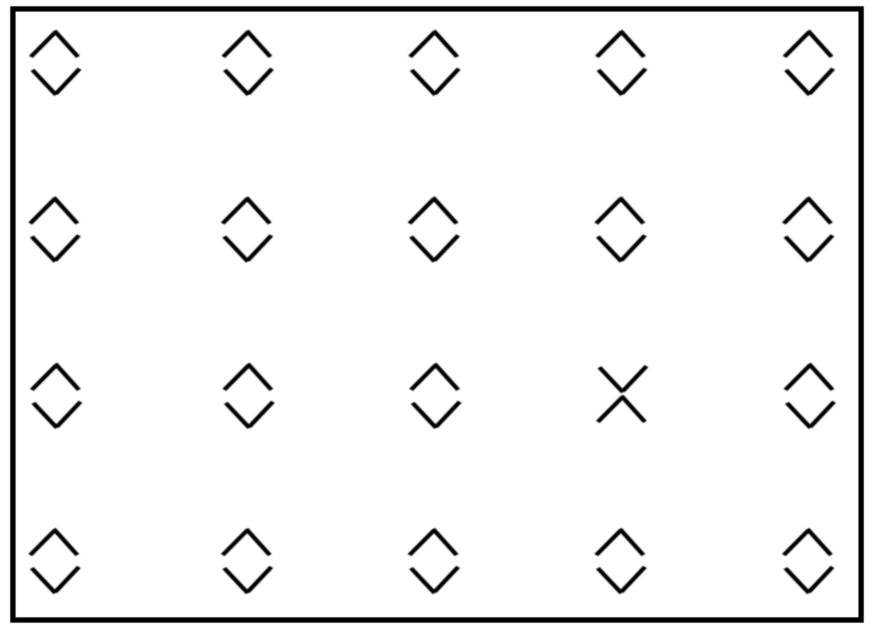
$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, I_1, I_2]}{\log(1/\epsilon)} = \alpha(I_1, I_2)$$



Fix error and characterise the time to find odd arm as a function of error





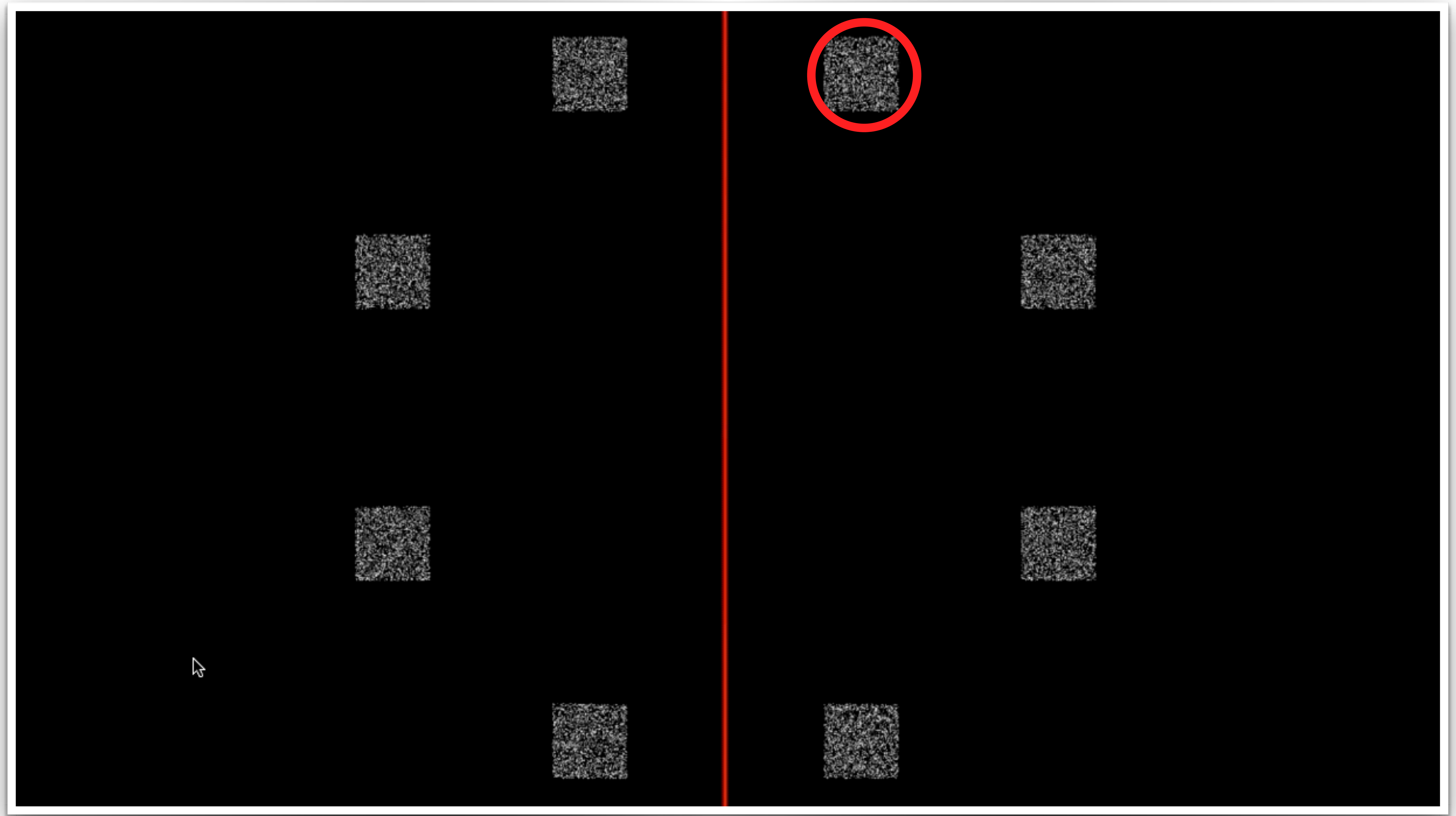


1. Vaidhiyan, N. K., Arun, S. P., & Sundaresan, R. (2017). Neural Dissimilarity Indices that Predict Oddball Detection in Behaviour. *IEEE Transactions on Information Theory*, *63*(8), 4778-4796.

2. Vaidhiyan, N. K., & Sundaresan, R. (2017). Learning to Detect an Oddball Target. *IEEE Transactions on Information Theory*, *64*(2), 831-852.

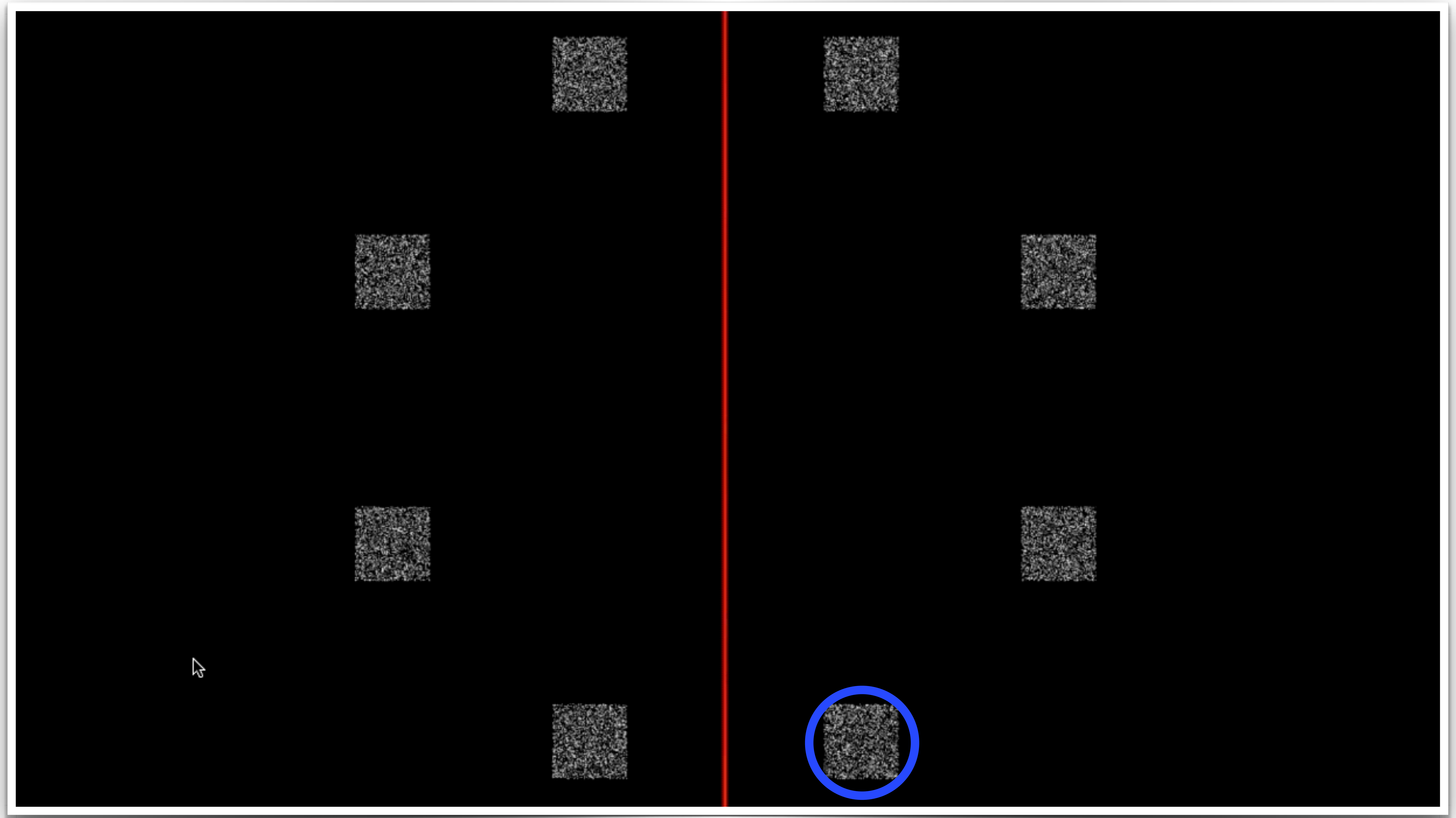
From Static Images to Movies

Find the Odd Movie — 1



Odd movie

Find the Odd Movie — 2



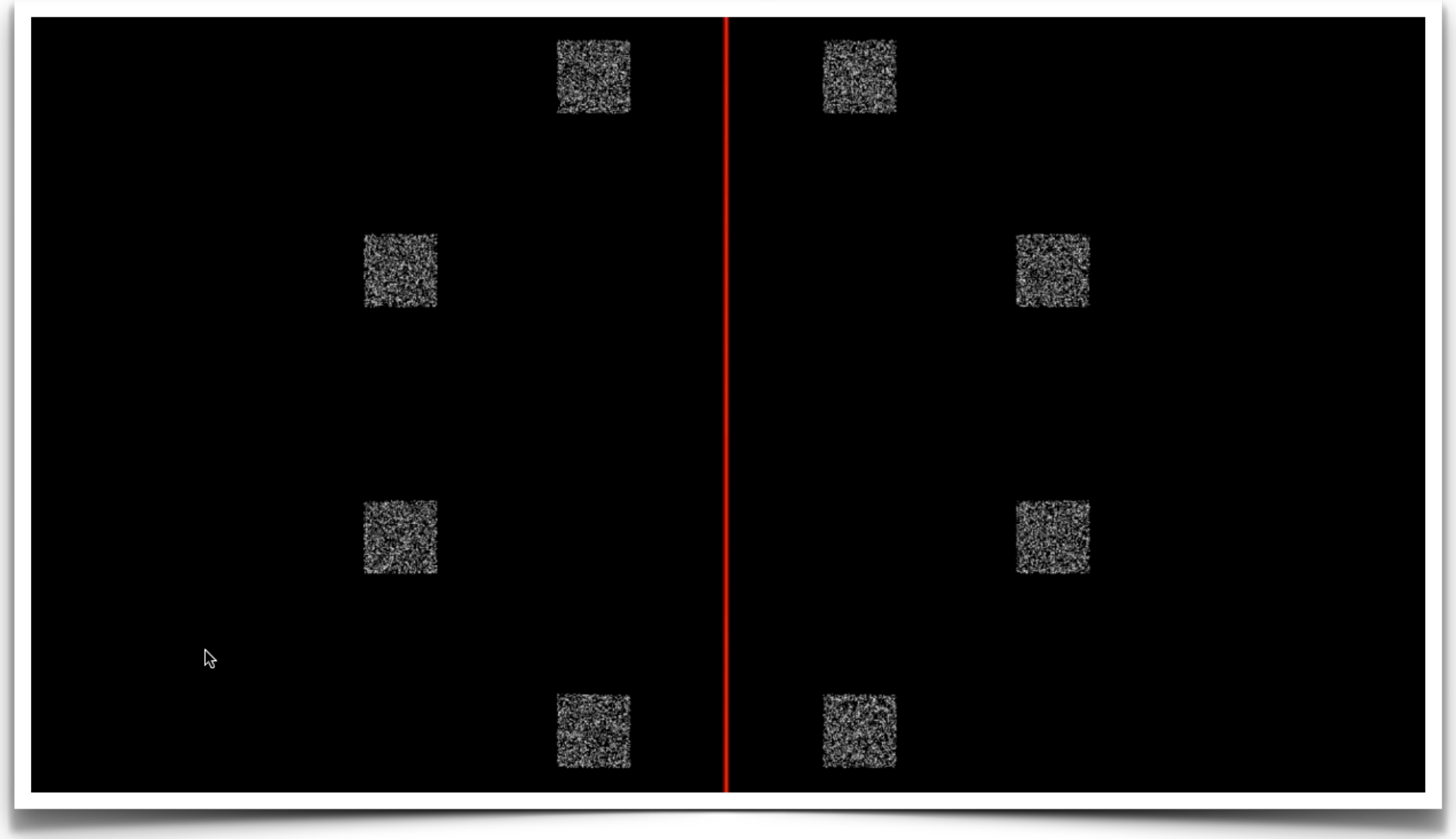
Odd movie

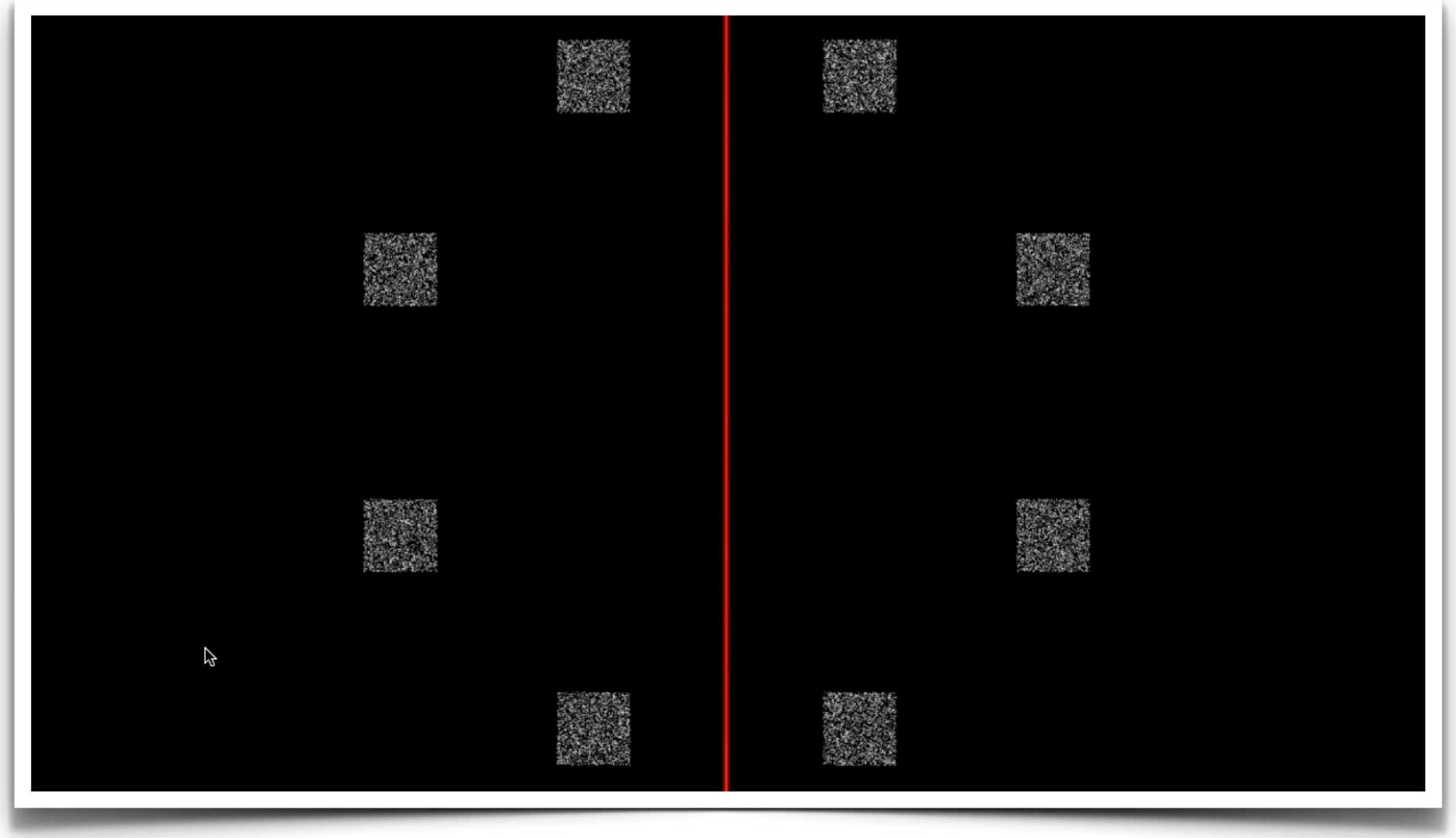
Finding the Odd Movie

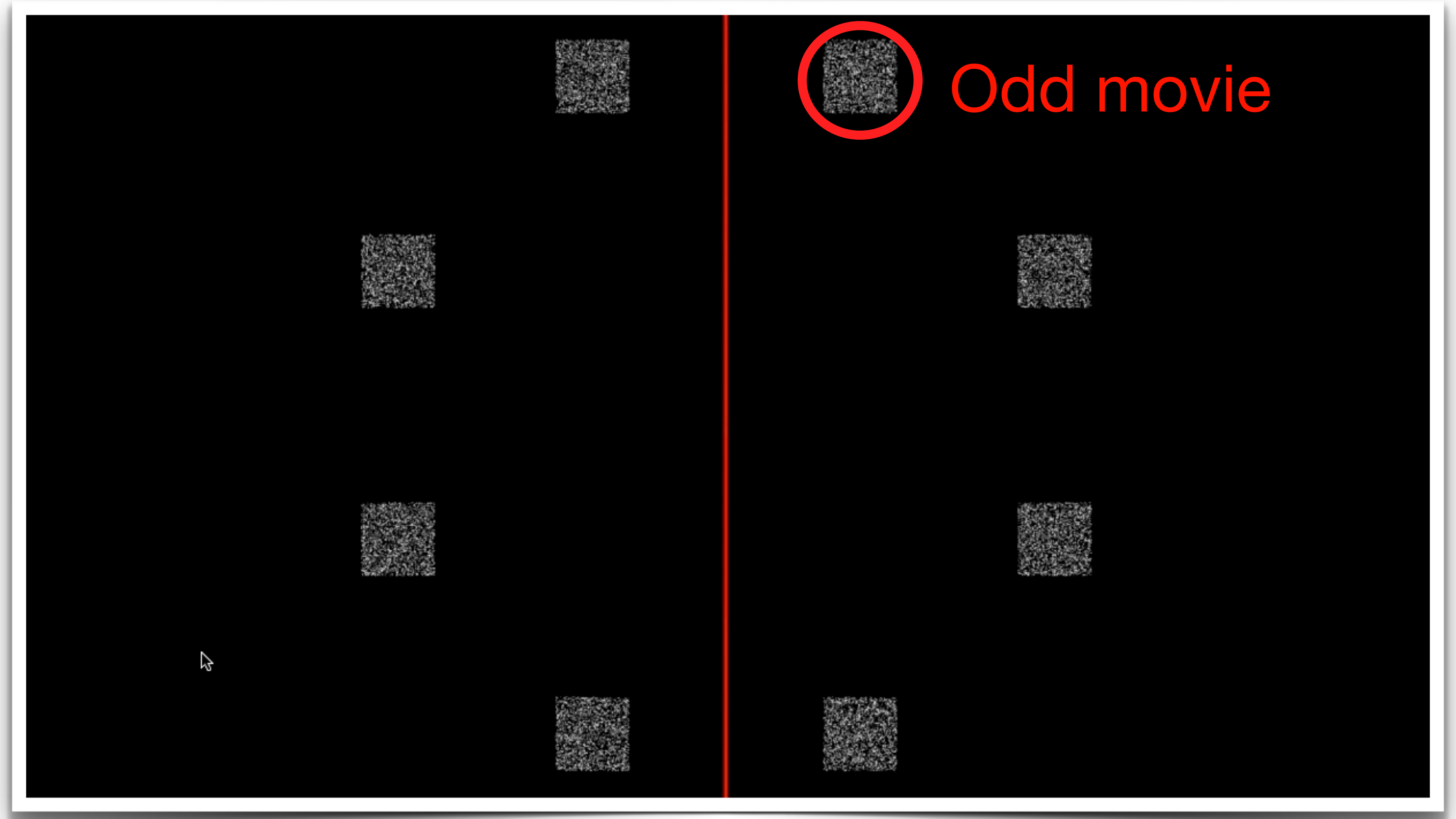
- Time to find the odd movie depends on the drifts of the movies
- The "closer" the drifts of the odd movie and the non-odd movies are, the longer it takes to find the odd movie
- Given movies with drifts d_1 and d_2, can we say

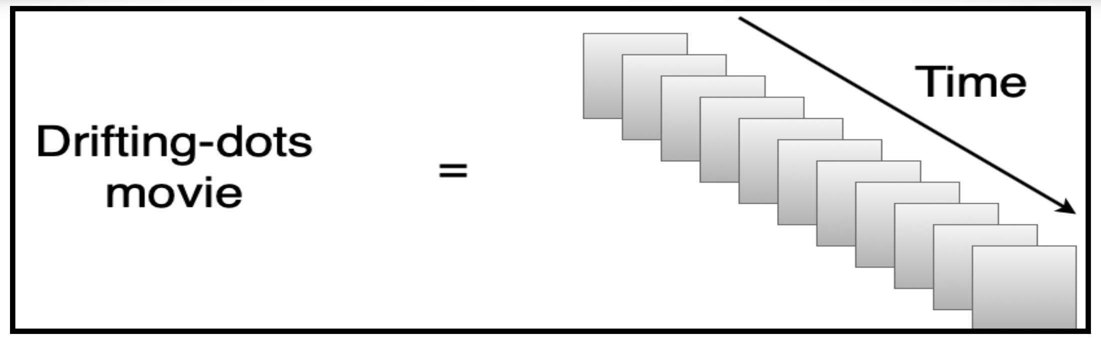
$$\inf_{\pi \in \Pi(\epsilon)} E[\tau(\pi) \,|\, d_1, d_2] \approx \alpha(d_1, d_2) \cdot \left(\log \frac{1}{\epsilon}\right)?$$

This talk: a detailed analysis of the above question



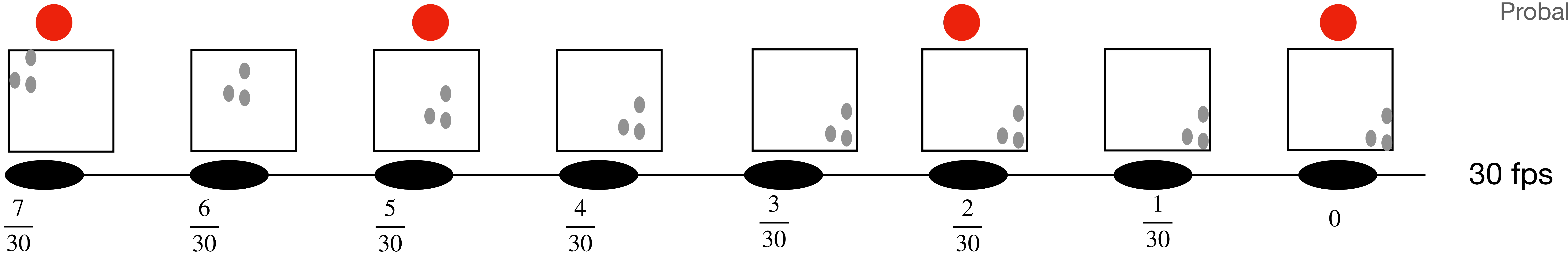






Odd Movie Experiments	Multi-Armed Bandits
Movie	Arm
Frame	Observation
Positions of dots in successive frames related	Observations form a Markov process
One movie is observed at a time	One arm is selected at a time
Unobserved movies continue to play	Unobserved arms continue to evolve (restless arms)
Drift of one of the movies is different	Markov law (TPM) of one of the arms is different

TPM: Transition Probability Matrix



The Odd Restless Markov Arm Problem

- A multi-armed bandit with $K \geq 3$ arms
- Each arm is a time homogeneous and ergodic Markov process
- Markov processes evolve on a common, finite state space
- The TPM of one of the arms (odd arm) is P_1; TPM of rest of the arms is P_2
- Arms are restless
- TPMs may be known beforehand or unknown

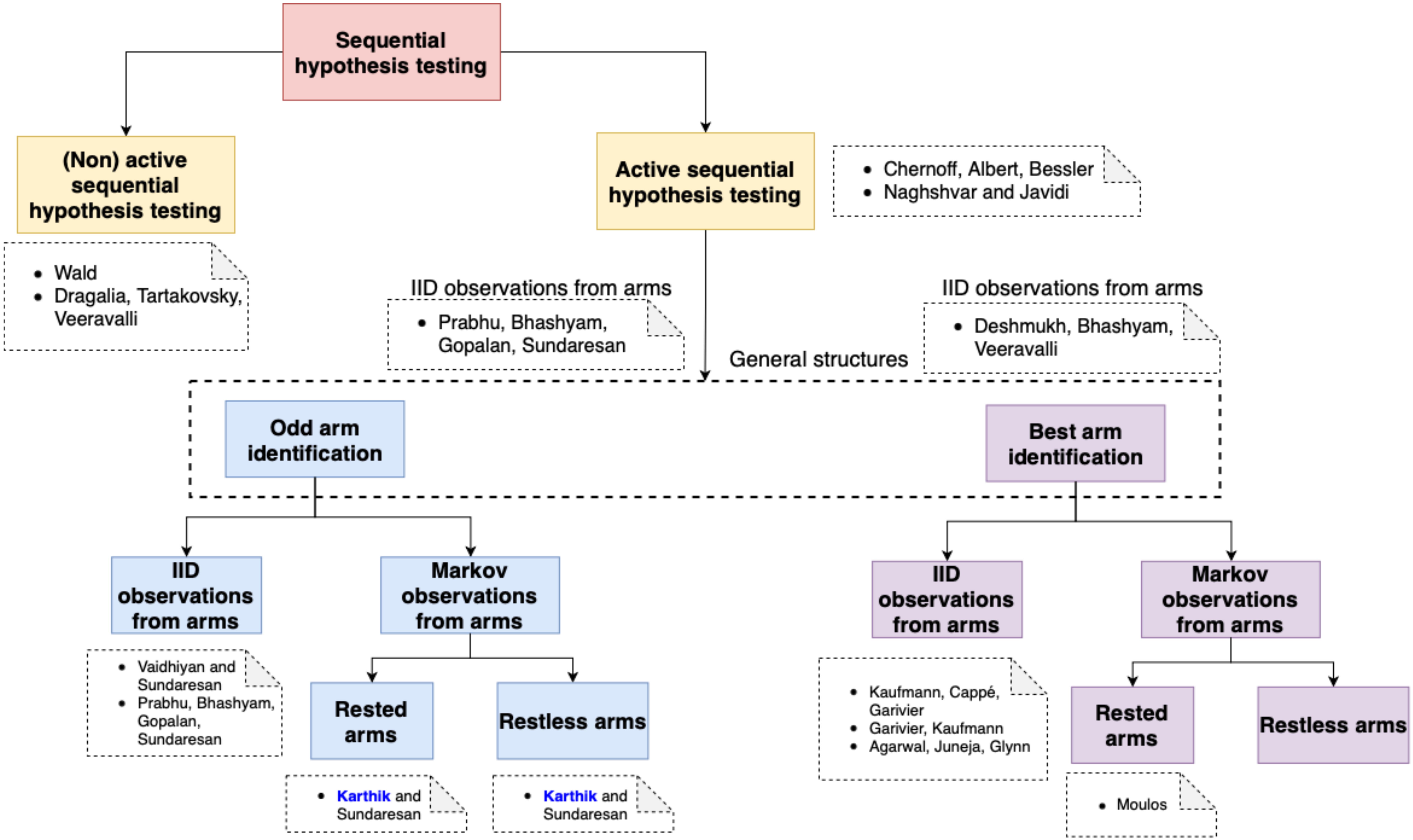
Characterise: $$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, P_1, P_2]}{\log(1/\epsilon)}$$

The Odd Rested Markov Arm Problem

- A multi-armed bandit with $K \geq 3$ arms
- Each arm is a time homogeneous and ergodic Markov process
- Markov processes evolve on a common, finite state space
- The TPM of one of the arms (odd arm) is P_1; TPM of rest of the arms is P_2
- Arms are rested
- TPMs may be known beforehand or unknown

Simpler to analyse; first step before analysing the more difficult setting of restless arms

Putting Our Work in Perspective — Optimal Stopping



Part 1: Rested Arms

P. N. Karthik and Rajesh Sundaresan, “Learning to Detect an Odd Markov Arm”, IEEE Transactions on Information Theory, 66(7), 4324-4348.

The Odd Rested Markov Arm Problem

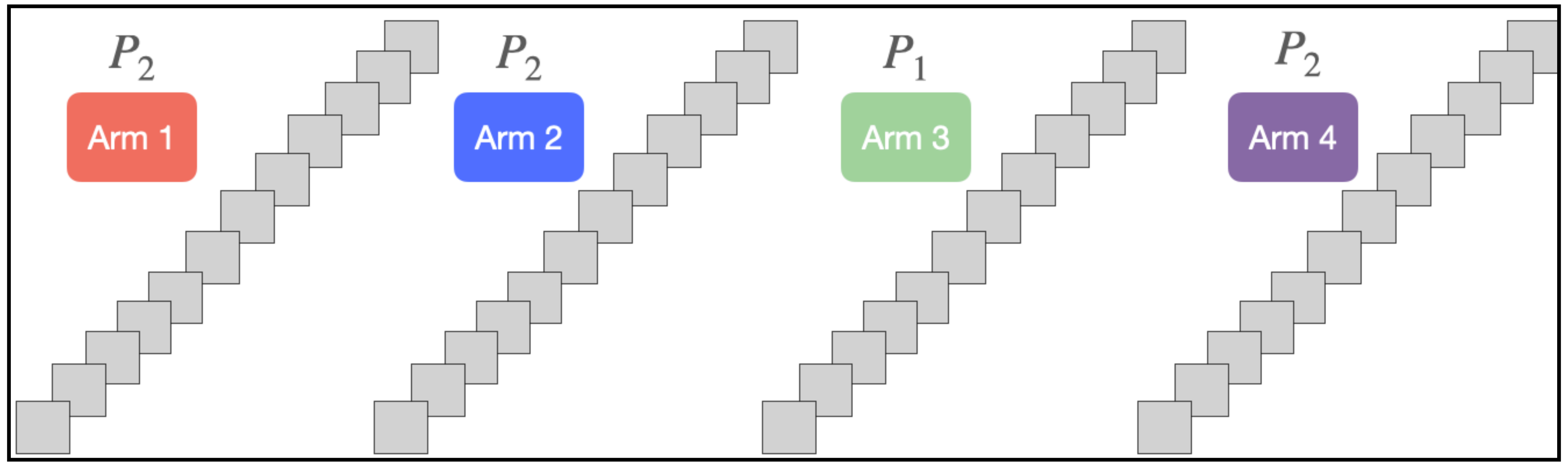
- A multi-armed bandit with $K \geq 3$ arms
- Each arm is a time homogeneous and ergodic Markov process
- Markov processes evolve on a common, finite state space
- The TPM of one of the arms (odd arm) is P_1; TPM of rest of the arms is P_2
- Arms are rested
- TPMs unknown (learning)

Our Contributions

- Let $C = (h, P_1, P_2)$ be a problem instance
- Lower bound:

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} \geq \frac{1}{D^*(h, P_1, P_2)}$$

- Policy — matching upper bound as $\epsilon \downarrow 0$

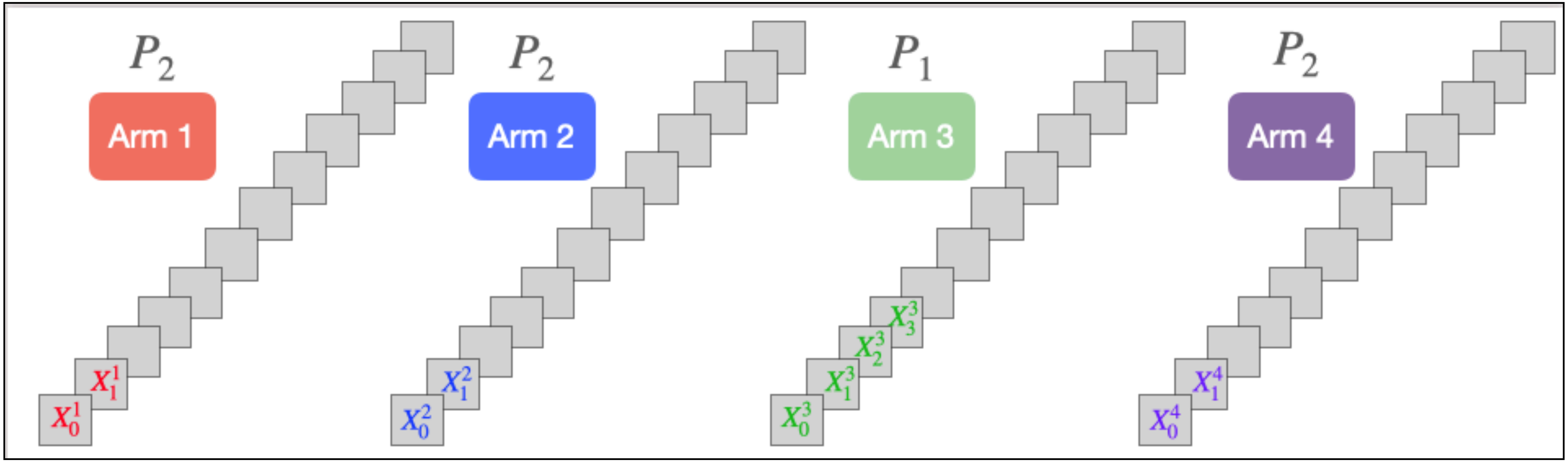


$$\boxed{\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} = \frac{1}{D^*(h, P_1, P_2)}}$$

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

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

X_t^a : tth observation from arm a

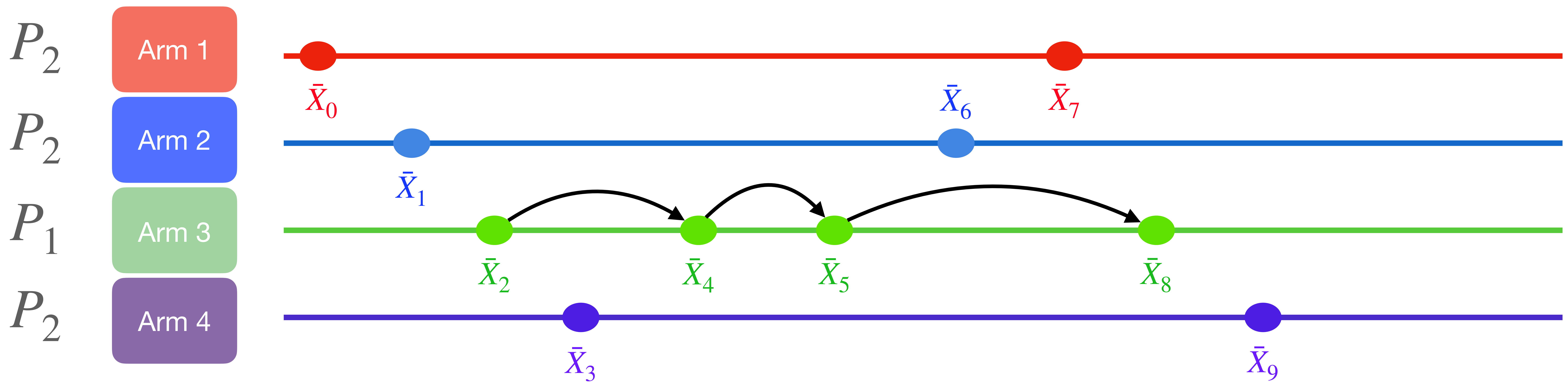


Arm a — sampled $N_a(n)$ times up to time n

$$A_0, \ldots, A_n, \underbrace{X_0^1, \ldots, X^1_{N_1(n)-1}}_{\text{arm } 1}, \ldots, \underbrace{X_0^K, \ldots, X^K_{N_K(n)-1}}_{\text{arm } K}$$

$$\boxed{C = (h, P_1, P_2)}$$

$$Z_C(n) = \log P(A_0, \ldots, A_n, X_0^1, \ldots, X^1_{N_1(n)-1}, \ldots, X_0^K, \ldots, X^K_{N_K(n)-1} \mid C)$$



$C = (h, P_1, P_2)$

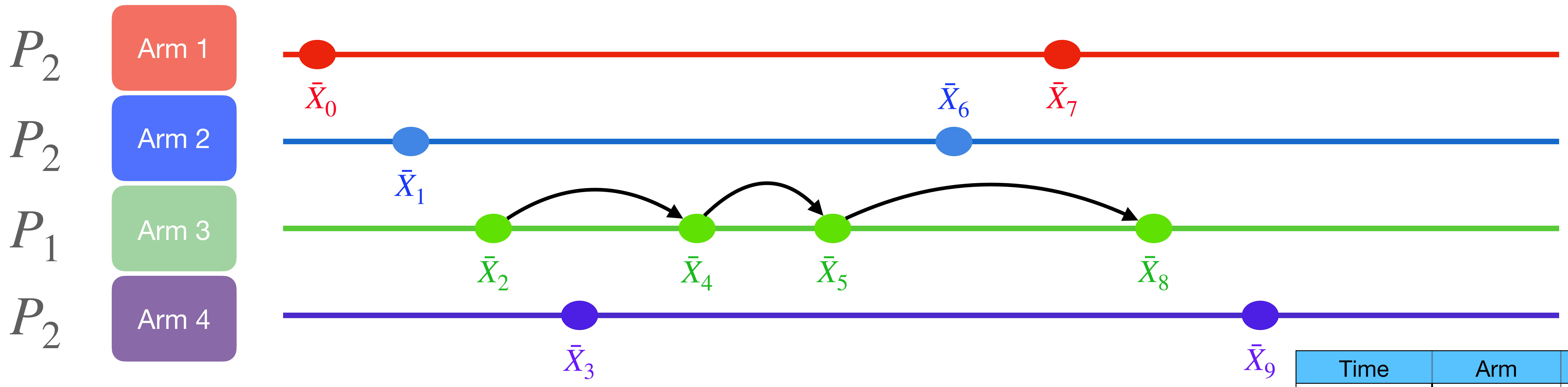
$$Z_C(n) = \sum_{i,j \in \mathcal{S}} N_h(n,i,j)\ \log P_1(j|i) + \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n,i,j)\ \log P_2(j|i) + \log P_C(A_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t | A_0, \ldots, A_{t-1}, \bar{X}_0, \ldots, \bar{X}_{t-1})$$

$C' = (h', P'_1, P'_2)$ $h' \neq h$

$$Z_{C'}(n) = \sum_{i,j \in \mathcal{S}} N_{h'}(n,i,j)\ \log P'_1(j|i) + \sum_{a \neq h'} \sum_{i,j \in \mathcal{S}} N_a(n,i,j)\ \log P'_2(j|i) + \log P_{C'}(A_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_{C'}(A_t | A_0, \ldots, A_{t-1}, \bar{X}_0, \ldots, \bar{X}_{t-1})$$

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

Converse: Key Ideas



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

$$\#\text{ transitions from } i = \#\text{ transitions to } i \pm 1$$

$$\lim_{n\to\infty} \frac{\#\text{ transitions from } i}{n} = \lim_{n\to\infty} \frac{\#\text{ transitions to } i}{n}$$

Converse: Key Ideas

$$d(\epsilon, 1-\epsilon) \leq E[Z_C(\tau(\pi)) - Z_{C'}(\tau(\pi)) \mid C] \lesssim E[\tau(\pi) \mid C] \cdot D^*(h, P_1, P_2)$$

Data processing inequality

Information theoretic bottleneck: maximum discrimination per unit time

$$\inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} \gtrsim \frac{1}{D^*(h, P_1, P_2)}$$

$\pi \in \Pi(\epsilon)$

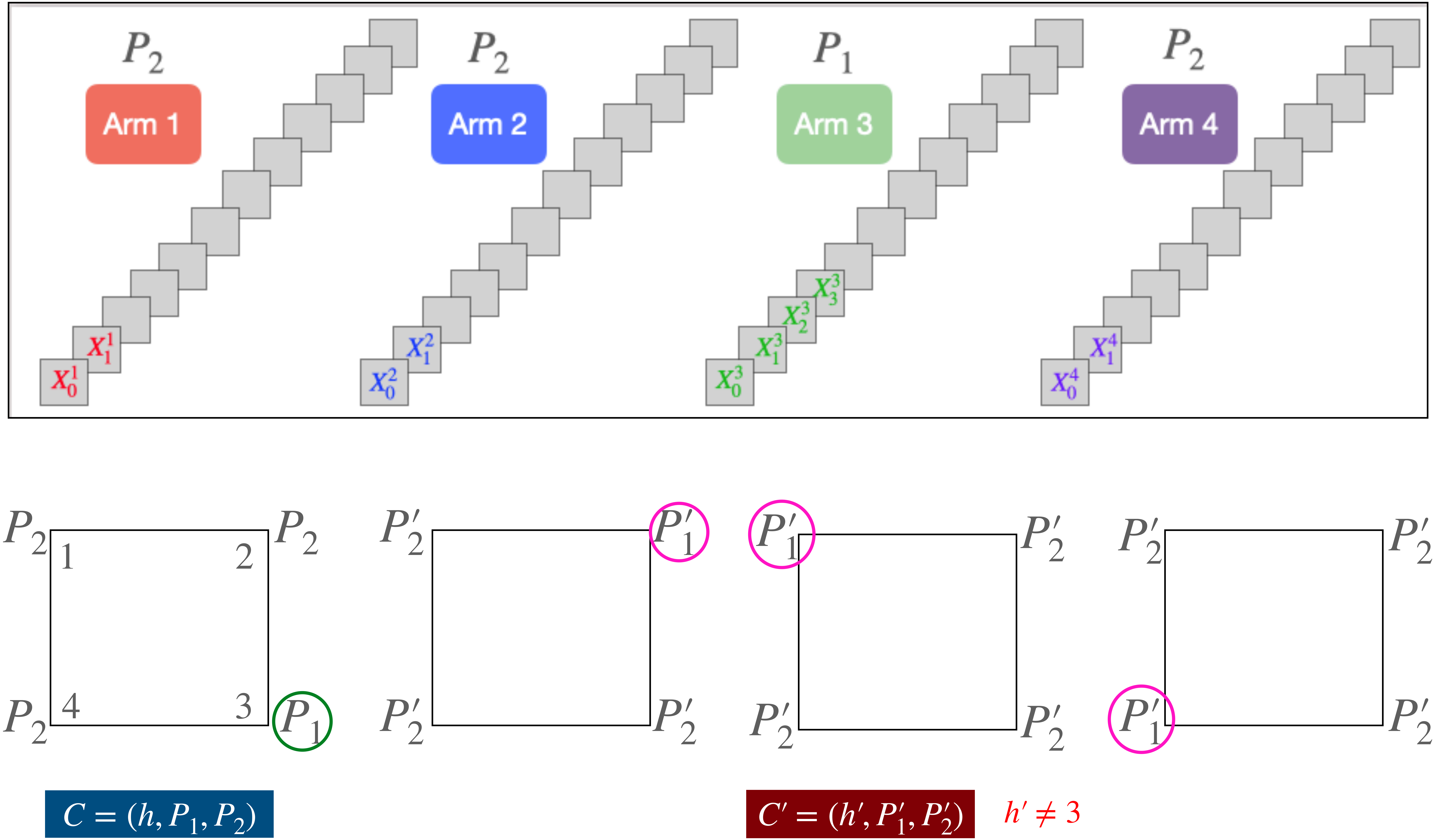
Configuration	Decision = h	Decision = h'	Others
$C = (h, P_1, P_2)$	$\geq 1-\epsilon$	$\leq \epsilon$	$\leq \epsilon$
$C' = (h', P_1', P_2')$	$\leq \epsilon$	$\geq 1-\epsilon$	$\leq \epsilon$

$C = (h, P_1, P_2)$

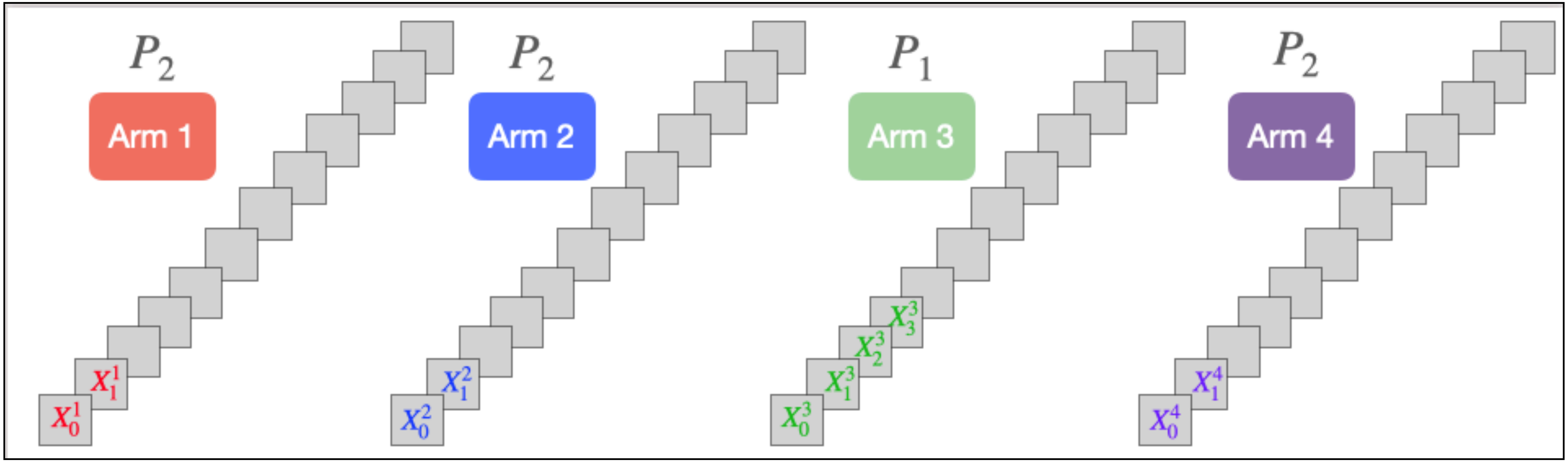
$$Z_C(n) = \sum_{i,j \in \mathcal{S}} N_h(n,i,j) \log P_1(j \mid i) + \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n,i,j) \log P_2(j \mid i) + \log P_C(A_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t \mid A_0, \ldots, A_{t-1}, \bar{X}_0, \ldots, \bar{X}_{t-1})$$

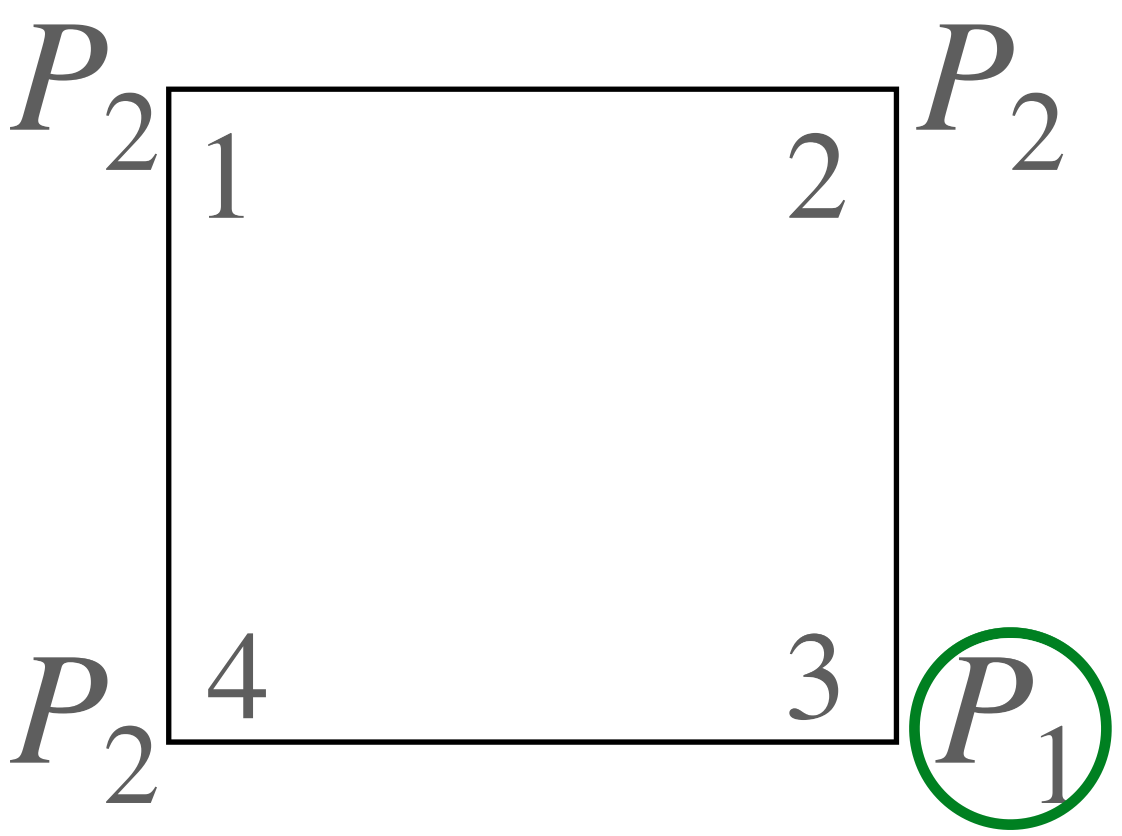
$C' = (h', P_1', P_2')$ $h' \neq h$

$$Z_{C'}(n) = \sum_{i,j \in \mathcal{S}} N_{h'}(n,i,j) \log P_1'(j \mid i) + \sum_{a \neq h'} \sum_{i,j \in \mathcal{S}} N_a(n,i,j) \log P_2'(j \mid i) + \log P_{C'}(A_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_{C'}(A_t \mid A_0, \ldots, A_{t-1}, \bar{X}_0, \ldots, \bar{X}_{t-1})$$



P_2
Arm 1
P_2
Arm 2
P_1
Arm 3
P_2
Arm 4
X_0^1 X_1^1
X_0^2 X_1^2
X_0^3 X_1^3 X_2^3 X_3^3
X_0^4 X_1^4
P_2 1 2 P_2
P_2 4 3 P_1
P'_2 P'_1 P'_2 P'_2
P'_1 P'_2 P'_2 P'_2
P'_2 P'_2 P'_1 P'_2
$C = (h, P_1, P_2)$
$C' = (h', P'_1, P'_2)$
$h' \neq 3$





$C = (h, P_1, P_2)$

Nearest alternative:

$P_1' = P_2$

$P_2' =$ convex combination of P_1 and P_2

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

Achievability

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

$$(h, P_1, P_2) \mapsto \lambda^*_{h, P_1, P_2}$$

continuous
(Berge's maximum theorem)

$$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n))$$

$$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n)) \approx (h, P_1, P_2)$$

$$\lambda^*_{\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),1}(n)} \approx \lambda^*_{h, P_1, P_2}$$

$$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n))$$

$$\hat{h}(n) \in \arg\max_{h} \min_{h' \neq h} M_{hh'}(n)$$

$$\hat{P}_{\hat{h}(n),1}(n)(j\,|\,i) = \frac{N_{\hat{h}(n)}(n,i,j)}{\sum_{j} N_{\hat{h}(n)}(n,i,j)}$$

$$\mathscr{D}(P_1, P_2) = \Gamma(P_1) \cdot \Gamma(P_2)$$

Prior $\Gamma(\cdot)$ on the space of all TPMs:

Pick each row of a TPM independently according to uniform distribution on the probability simplex.

$$\text{average likelihood}_h(n) = \int_{P_1,P_2} \exp(Z_C(n))\ \mathscr{D}(P_1,P_2)\ dP_1\ dP_2$$

$$\text{maximum likelihood}_h(n) = \max_{P_1,P_2} Z_C(n)$$

$$\hat{P}_{\hat{h}(n),2}(n)(j\,|\,i) = \frac{\sum_{a \neq \hat{h}(n)} N_a(n,i,j)}{\sum_{a \neq \hat{h}(n)} \sum_{j} N_a(n,i,j)}$$

$$M_{hh'}(n) = \frac{\text{average likelihood}_h(n)}{\text{maximum likelihood}_{h'}(n)}$$

$$Z_C(n) = \sum_{i,j \in \mathcal{S}} N_h(n,i,j)\ \log P_1(j\,|\,i) + \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n,i,j)\ \log P_2(j\,|\,i) + \log P_C(A_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t\,|\,A_0,\ldots,A_{t-1},\bar{X}_0,\ldots,\bar{X}_{t-1})$$

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

$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n)) \approx (h, P_1, P_2)$ $\qquad$ $\lambda^*_{\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),1}(n)} \approx \lambda^*_{h,P_1,P_2}$

- Select each arm once ($n = 0,\ldots,K-1$)
- For $n \geq K$, repeat the following until stoppage:
  - Estimate $\hat{h}(n)$
  - If $\min_{h' \neq \hat{h}(n)} M_{\hat{h}(n),h'}(n) \geq \log((K-1)L)$, stop and declare $\hat{h}(n)$ as the odd arm
  - Else, toss a coin with $\Pr(\text{heads}) = \delta$
    - If coin lands heads, sample an arm uniformly randomly
    - If coin lands tails, sample according to $\lambda^*_{\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n)}$

Certainty Equivalence

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

$$(h, P_1, P_2) \mapsto \lambda^*_{h,P_1,P_2}$$

$$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n))$$

$$\hat{h}(n) \in \arg\max_{h} \min_{h' \neq h} M_{hh'}(n)$$

$$\hat{P}_{\hat{h}(n),1}(n)(j\,|\,i) = \frac{N_{\hat{h}(n)}(n,i,j)}{\sum_j N_{\hat{h}(n)}(n,i,j)}$$

$$\hat{P}_{\hat{h}(n),2}(n)(j\,|\,i) = \frac{\sum_{a \neq \hat{h}(n)} N_a(n,i,j)}{\sum_{a \neq \hat{h}(n)} \sum_j N_a(n,i,j)}$$

Why not Sample the Arms Repeatedly?

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

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

$$(h, P_1, P_2) \mapsto \lambda^*_{h,P_1,P_2}$$

$$\hat{h}(n) \in \arg\max_{h} \min_{h' \neq h} M_{hh'}(n)$$

$$\hat{P}_{\hat{h}(n),1}(n)(j \,|\, i) = \frac{N_{\hat{h}(n)}(n,i,j)}{\sum_j N_{\hat{h}(n)}(n,i,j)}$$

$$\hat{P}_{\hat{h}(n),2}(n)(j \,|\, i) = \frac{\sum_{a \neq \hat{h}(n)} N_a(n,i,j)}{\sum_{a \neq \hat{h}(n)} \sum_j N_a(n,i,j)}$$

$$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n)) \approx (h, P_1, P_2)$$

$$\lambda^*_{\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),1}(n)} \approx \lambda^*_{h,P_1,P_2}$$

$$\Pi(\epsilon) = \{\pi : P_{\text{error}}(\pi) \leq \epsilon\}$$

Performance of $\pi^\star(L, \delta)$

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

- Stops in finite time w.p. 1
- $\hat{h}(n) = h$ for all n large, almost surely
- $(\hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n)) \longrightarrow (P_1, P_2)$ (ergodic theorem)
- If $L = 1/\epsilon,$ then $\pi^\star(L, \delta) \in \Pi(\epsilon)$
- Upper bound:

$$\limsup_{L \to \infty} \frac{E[\tau(\pi^\star(L, \delta)) \,|\, h, P_1, P_2]}{\log L} \le \frac{1}{D_\delta\,(h, P_1, P_2)}, \qquad D_\delta\,(h, P_1, P_2) \longrightarrow D^*(h, P_1, P_2) \quad \text{as} \quad \delta \downarrow 0$$

- Therefore,

$$\lim_{\delta \downarrow 0}\ \limsup_{L \to \infty} \frac{E[\tau(\pi^\star(L, \delta)) \,|\, h, P_1, P_2]}{\log L} \le \frac{1}{D^*(h, P_1, P_2)}$$

$$\boxed{\lim_{\epsilon \downarrow 0}\ \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)} = \frac{1}{D^*(h, P_1, P_2)}}$$

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

Part 2: Restless Arms with TPMs Known

P. N. Karthik and Rajesh Sundaresan, "Detecting an Odd Restless Markov Arm with a Trembling Hand", IEEE Transactions on Information Theory, 2021.

The Odd Restless Markov Arm Problem with Known TPMs

- A multi-armed bandit with $K \geq 3$ arms
- Each arm is a time homogeneous and ergodic Markov process
- Markov processes evolve on a common, finite state space
- The TPM of one of the arms (odd arm) is P_1; TPM of rest of the arms is P_2
- Arms are restless
- TPMs known beforehand

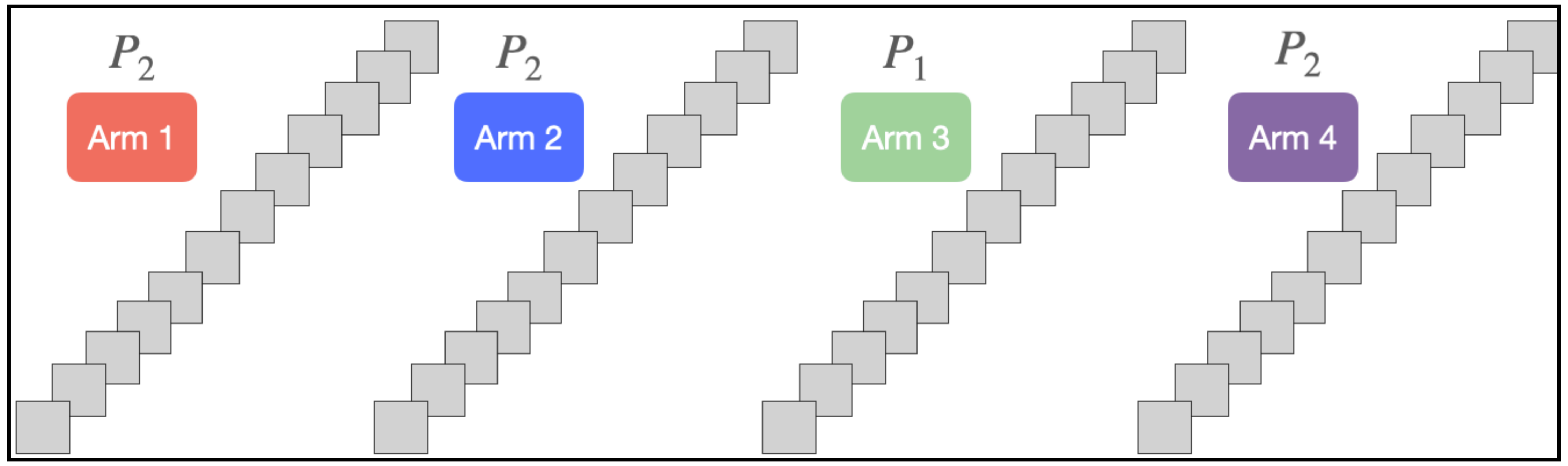
Characterise $\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)}$

Our Contributions

- Let $C = (h, P_1, P_2)$ be a problem instance
- Lower bound:

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

- Policy — matching upper bound as $\epsilon \downarrow 0$



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

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\, k_{hh'}(\underline{d}, \underline{i}, a)$$

Trembling Hand

- Often in visual search experiments, at each time t, the actual focus location (A_t) differs from the intended focus location (B_t) with small probability

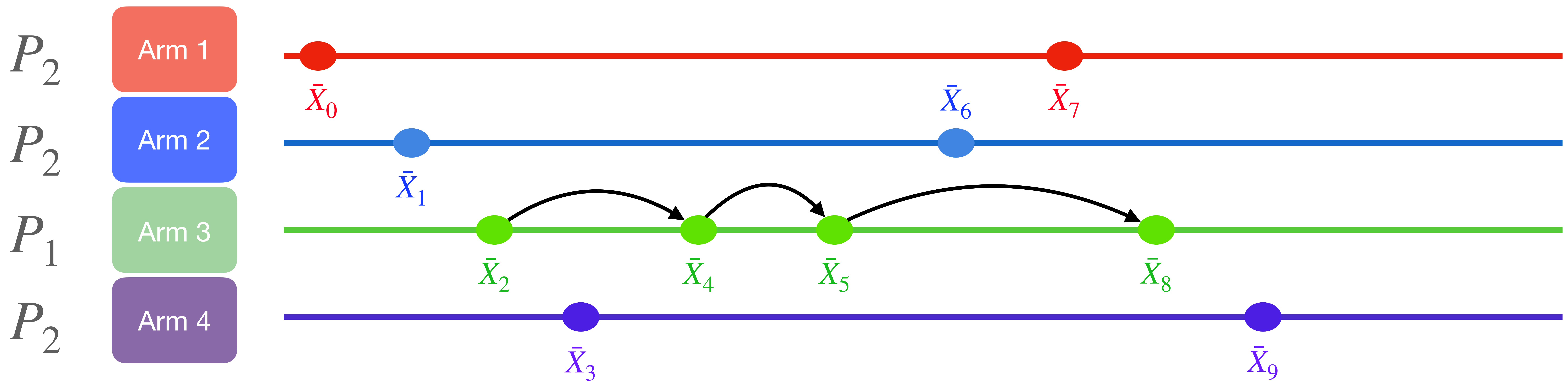
- This can be captured as a trembling hand:

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

- $\eta \in (0,1]$: trembling hand parameter

t	A_t	**Arm 1**		**Arm 2**		**Arm 3**		**Arm 4**	
		$d_1(t)$	$i_1(t)$	$d_2(t)$	$i_2(t)$	$d_3(t)$	$i_3(t)$	$d_4(t)$	$i_4(t)$
0	1								
1	2								
2	3								
3	4								
4	3	4	X_0^1	3	X_1^2	2	X_2^3	1	X_3^4
5	3	5	X_0^1	4	X_1^2	1	X_4^3	2	X_3^4
6	2	6	X_0^1	5	X_1^2	1	X_5^3	3	X_3^4
7	1	7	X_0^1	1	X_6^2	2	X_5^3	4	X_3^4
8	3	1	X_7^1	2	X_6^2	3	X_5^3	5	X_3^4
9		2	X_7^1	3	X_6^2	1	X_8^3	6	X_3^4

X_t^a : observation from arm a at time t $d_a(t)$: delay of arm a at time t $i_a(t)$: last observed state of arm a at time t

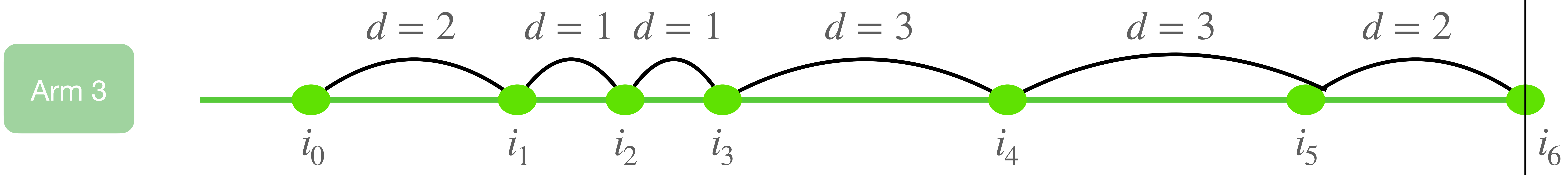


Time	Arm	Obs.
0	1	X_0^1
1	2	X_1^2
2	3	X_2^3
3	4	X_3^4
4	3	X_4^3
5	3	X_5^3
6	2	X_6^2
7	1	X_7^1
8	3	X_8^3
9	4	X_9^4

$C = (h, P_1, P_2)$

$$Z_C(n) = \sum_d \sum_{i,j \in \mathcal{S}} N_h(n,d,i,j) \; \log P_1^d(j \mid i) + \sum_d \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n,d,i,j) \; \log P_2^d(j \mid i)$$
$$+ \log P_C(A_0, B_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t, B_t \mid B_0^{t-1}, A_0^{t-1}, \bar{X}_0^{t-1})$$

Arm\Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1															
2															
3															
4															



$$\sum_j N_a(n, d, i, j) = N_a(n, d, i)$$

$n = 14$

$$(a - 1) + \sum_{i \in \mathcal{S}} \sum_{d=1}^{\infty} d \cdot N_a(n, d, i) = n$$

Arm\Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1															
2															
3															
4															

$$\lim_{n \to \infty} \sum_{i \in \mathcal{S}} \sum_{d=1}^{\infty} d \cdot \frac{N_a(n, d, i)}{n} = 1$$

$$E[Z_h(\tau(\pi)) - Z_{h'}(\tau(\pi)) \mid C] \lesssim E[\tau(\pi) \mid C] \cdot R_1^*(P_1, P_2)$$

Information theoretic bottleneck:
maximum discrimination per unit time

$C = (h, P_1, P_2)$

$$R_1^*(P_1, P_2) = \sup_{\kappa} \min_{h' \neq h} \sum_{a=1}^{K} \sum_{d=1}^{\infty} \sum_{i \in \mathcal{S}} \kappa(d, i, a)\, D((P_h^a)^d(\,\cdot \mid i) \| (P_{h'}^a)^d(\,\cdot \mid i))$$

subject to

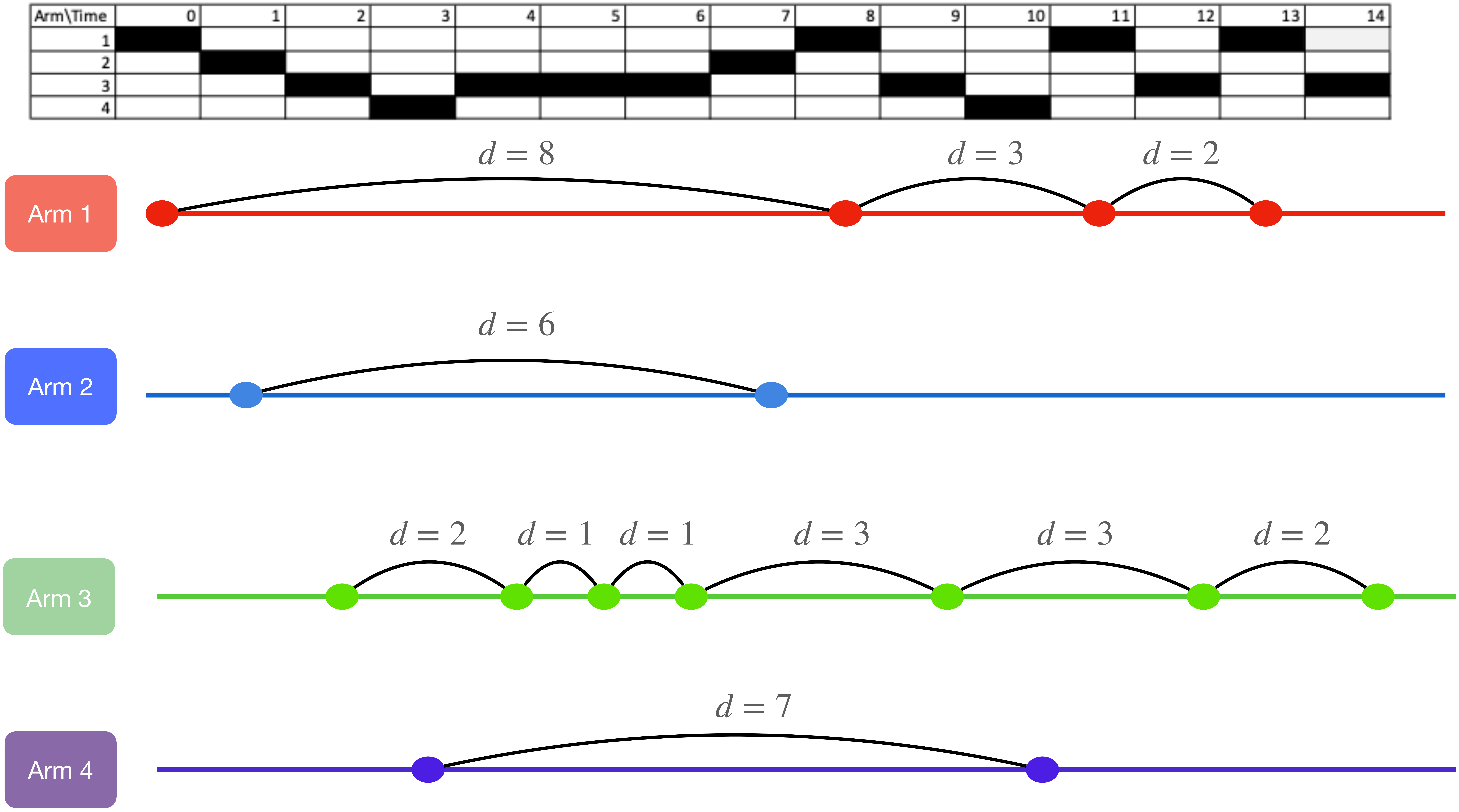
$$\sum_{i \in \mathcal{S}} \sum_{d=1}^{\infty} d\ \kappa(d, i, a) = 1 \quad \text{for all } a,$$

$$\sum_{d=1}^{\infty} \sum_{i \in \mathcal{S}} \sum_{a=1}^{K} \kappa(d, i, a) = 1,$$

$$\kappa(d, i, a) \geq 0 \quad \text{for all } a,\ d \in \{1, 2, \ldots\},\ i \in \mathcal{S}$$

$$Z_h(n) = \sum_{d} \sum_{i,j \in \mathcal{S}} N_h(n, d, i, j)\ \log P_1^d(j \mid i) + \sum_{d} \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n, d, i, j)\ \log P_2^d(j \mid i)$$

$$+ \log P_C(A_0, B_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t, B_t \mid B_0^{t-1}, A_0^{t-1}, \bar{X}_0^{t-1})$$



Arm\Time
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
1
2
3
4
d = 8
d = 3
d = 2
Arm 1
d = 6
Arm 2
d = 2
d = 1
d = 1
d = 3
d = 3
d = 2
Arm 3
d = 7
Arm 4

Delays and Last Observed States

- $\underline{d}(t) = (d_1(t), \ldots, d_K(t))$ $\qquad$ $\underline{i}(t) = (i_1(t), \ldots, i_K(t))$
- $(B_0, A_0, X_0^{A_0}, \ldots, B_{t-1}, A_{t-1}, X_{t-1}^{A_{t-1}}) \equiv (B_0, \ldots, B_{t-1}, \{\underline{d}(s), \underline{i}(s) : K \leq s \leq t\})$
- $\{(\underline{d}(t), \underline{i}(t)) : t \geq K\}$ is a controlled Markov process with controls $\{B_t : t \geq 0\}$

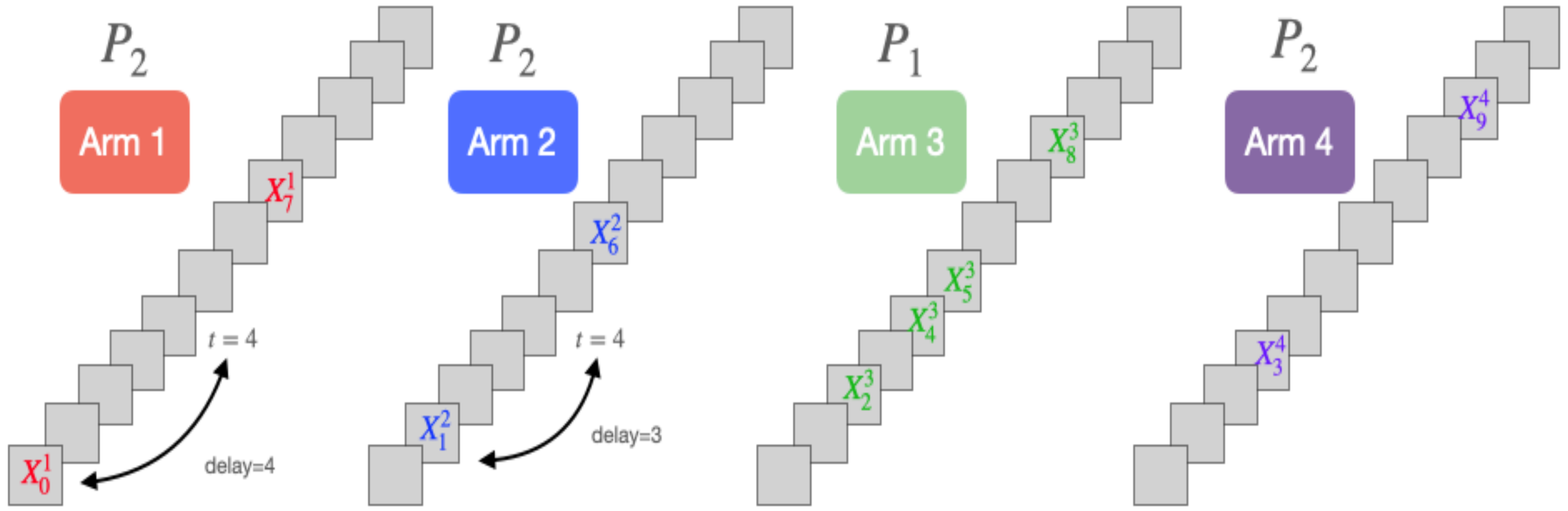
$$P(\underline{d}(t+1), \underline{i}(t+1) \,|\, B_0, \ldots, B_t, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = P(\underline{d}(t+1), \underline{i}(t+1) \,|\, B_t, (\underline{d}(t), \underline{i}(t)))$$

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

Markov Decision Problem (MDP)

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

State space	$\mathbb{S} = \{(\underline{d}, \underline{i})\}$
Action space	$\{1, \ldots, K\}$
State at time t	$(\underline{d}(t), \underline{i}(t))$
Action at time t	B_t



t	A_t	Arm 1		Arm 2		Arm 3		Arm 4	
		$d_1(t)$	$i_1(t)$	$d_2(t)$	$i_2(t)$	$d_3(t)$	$i_3(t)$	$d_4(t)$	$i_4(t)$
0	1								
1	2								
2	3								
3	4								
4	3	4	X_0^1	3	X_1^2	2	X_2^3	1	X_3^4
5	3	5	X_0^1	4	X_1^2	1	X_4^3	2	X_3^4
6	2	6	X_0^1	5	X_1^2	1	X_5^3	3	X_3^4
7	1	7	X_0^1	1	X_6^2	2	X_5^3	4	X_3^4
8	3	1	X_7^1	2	X_6^2	3	X_5^3	5	X_3^4
9		2	X_7^1	3	X_6^2	1	X_8^3	6	X_3^4

X_t^a : observation from arm a at time t $\quad d_a(t)$: delay of arm a at time t $\quad i_a(t)$: last observed state of arm a at time t

Characterise $\displaystyle \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)}$

Markov Decision Problem (MDP)

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

MDP Transition Probabilities

$$\underline{d}(t) = \underline{d} = (4,3,2,1) \qquad \underline{i}(t) = \underline{i} = (i_1, i_2, i_3, i_4)$$

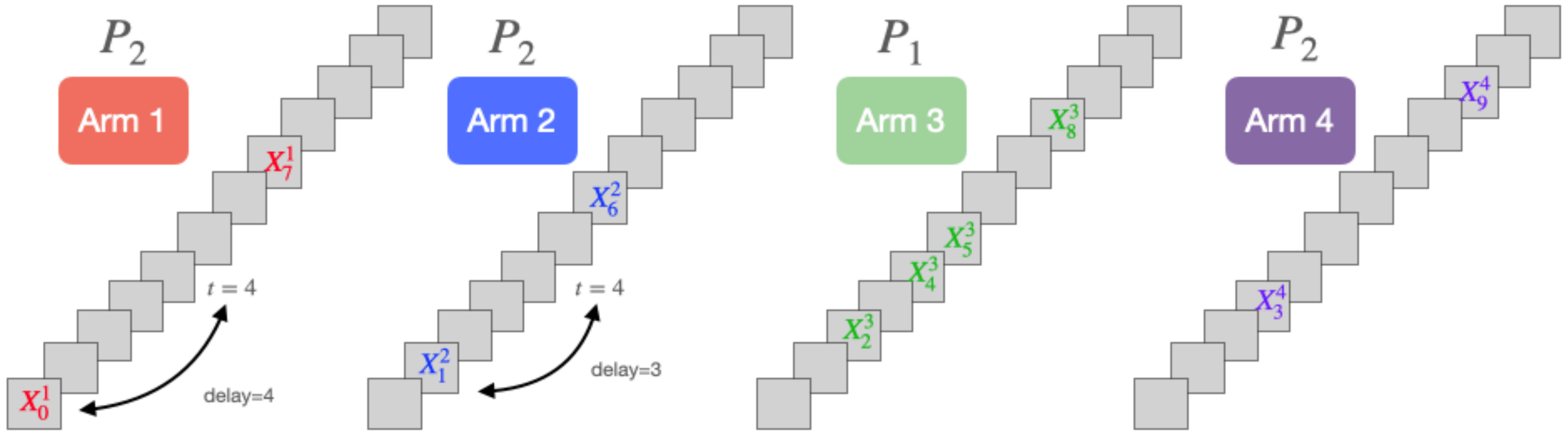
$$B_t = b \qquad A_t = 1$$

$$P(A_t = 1 \,|\, B_t = b) = \frac{\eta}{K} + (1-\eta)\,\mathbb{1}_{\{b=1\}}$$

$$\underline{d}(t+1) = \underline{d}^{'} = (1,4,3,2) \qquad \underline{i}(t+1) = \underline{i}^{'} = (X_t^1, i_2, i_3, i_4)$$

$$\underbrace{P(\underline{d}(t+1) = \underline{d}^{'}, \underline{i}(t+1) = \underline{i}^{'} \,|\, \underline{d}(t) = \underline{d}, \underline{i}(t) = \underline{i}, B_t = b)}_{Q(\underline{d}^{'}, \underline{i}^{'} \,|\, \underline{d}, \underline{i}, b)} = \left(\frac{\eta}{K} + (1-\eta)\,\mathbb{1}_{\{b=1\}}\right)(P_2)^4(X_t^1 \,|\, i_1)$$

Fourth power of the TPM P_2



t	A_t	Arm 1		Arm 2		Arm 3		Arm 4	
		$d_1(t)$	$i_1(t)$	$d_2(t)$	$i_2(t)$	$d_3(t)$	$i_3(t)$	$d_4(t)$	$i_4(t)$
0	1								
1	2								
2	3								
3	4								
4	3	4	X_0^1	3	X_1^2	2	X_2^3	1	X_3^4
5	3	5	X_0^1	4	X_1^2	1	X_4^3	2	X_3^4
6	2	6	X_0^1	5	X_1^2	1	X_5^3	3	X_3^4
7	1	7	X_0^1	1	X_6^2	2	X_5^3	4	X_3^4
8	3	1	X_7^1	2	X_6^2	3	X_5^3	5	X_3^4
9		2	X_7^1	3	X_6^2	1	X_8^3	6	X_3^4

X_t^a : observation from arm a at time t $\quad$ $d_a(t)$: delay of arm a at time t $\quad$ $i_a(t)$: last observed state of arm a at time t

Characterise $$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)}$$

Markov Decision Problem (MDP)

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

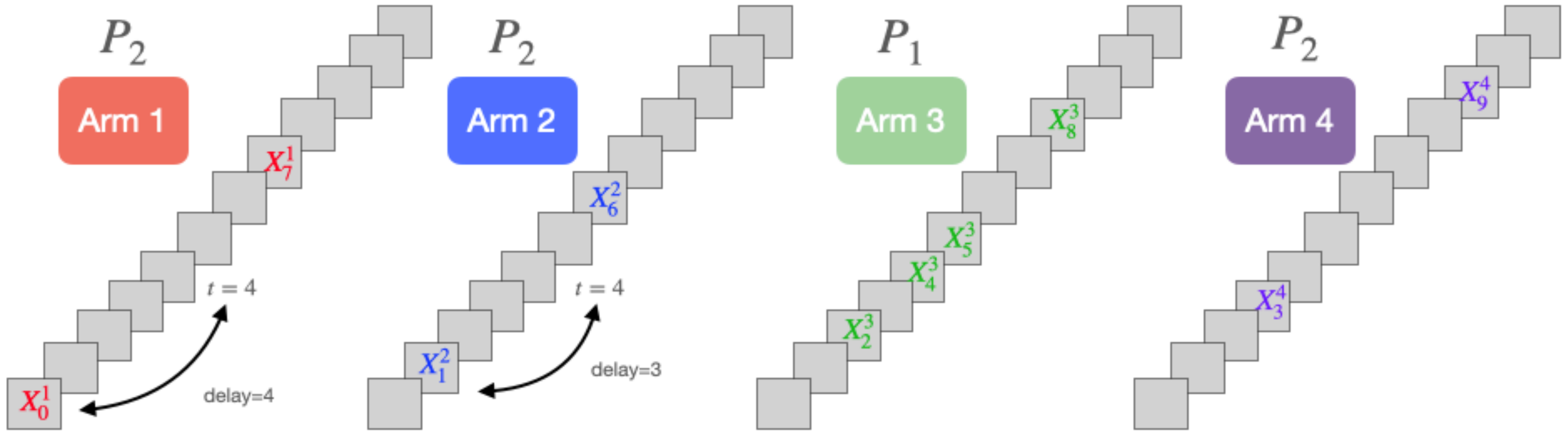
MDP Transition Probabilities

$$\underline{d}(t+1) = \underline{d}^{'} = (1,4,3,2) \qquad \underline{i}(t+1) = \underline{i}^{'} = (X_t^1, i_2, i_3, i_4)$$

$$B_{t+1} = b' \qquad A_{t+1} = 3$$

$$P(A_{t+1} = 3 \,|\, B_{t+1} = b') = \frac{\eta}{K} + (1-\eta)\,\mathbb{1}_{\{b'=3\}}$$

$$\underline{d}(t+2) = \underline{d}^{''} = (2,5,1,3) \qquad \underline{i}(t+2) = \underline{i}^{''} = (X_t^1, i_2, X_{t+1}^3, i_4)$$



t	A_t	Arm 1		Arm 2		Arm 3		Arm 4	
		$d_1(t)$	$i_1(t)$	$d_2(t)$	$i_2(t)$	$d_3(t)$	$i_3(t)$	$d_4(t)$	$i_4(t)$
0	1								
1	2								
2	3								
3	4								
4	3	4	X_0^1	3	X_1^2	2	X_2^3	1	X_3^4
5	3	5	X_0^1	4	X_1^2	1	X_4^3	2	X_3^4
6	2	6	X_0^1	5	X_1^2	1	X_5^3	3	X_3^4
7	1	7	X_0^1	1	X_6^2	2	X_5^3	4	X_3^4
8	3	1	X_7^1	2	X_6^2	3	X_5^3	5	X_3^4
9		2	X_7^1	3	X_6^2	1	X_8^3	6	X_3^4

X_t^a : observation from arm a at time t $\quad$ $d_a(t)$: delay of arm a at time t $\quad$ $i_a(t)$: last observed state of arm a at time t

$$\underbrace{P(\underline{d}(t+2) = \underline{d}^{''}, \underline{i}(t+2) = \underline{i}^{''} \,|\, \underline{d}(t+1) = \underline{d}^{'}, \underline{i}(t+1) = \underline{i}^{'}, B_{t+1} = b')}_{Q(\underline{d}^{''}, \underline{i}^{''} \,|\, \underline{d}^{'}, \underline{i}^{'}, b)} = \left(\frac{\eta}{K} + (1-\eta)\,\mathbb{1}_{\{b'=3\}}\right)(P_1)^3(X_{t+1}^3 \,|\, i_3)$$

Third power of the TPM P_1

Characterise $\displaystyle \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)}$

Data processing inequality

$$d(\epsilon, 1-\epsilon) \le E[Z_h(\tau(\pi)) - Z_{h'}(\tau(\pi)) \mid C] \lesssim E[\tau(\pi) \mid C] \cdot R^*(P_1, P_2)$$

Information theoretic bottleneck: maximum discrimination per unit time

$C = (h, P_1, P_2)$

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} \ge \frac{1}{R^*(P_1, P_2)}$$

$$R^*(P_1, P_2) = \sup_{\nu} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\ k_{hh'}(\underline{d}, \underline{i}, a)$$

$$k_{hh'}(\underline{d}, \underline{i}, a) = \begin{cases} D(P_1^{d_a}(\cdot \mid i_a) \| P_2^{d_a}(\cdot \mid i_a)), & a = h, \\ D(P_2^{d_a}(\cdot \mid i_a) \| P_1^{d_a}(\cdot \mid i_a)), & a = h', \\ 0, & a \neq h, h', \end{cases}$$

$\pi \in \Pi(\epsilon)$

Configuration	Decision = h	Decision = h'	Others
$C = (h, P_1, P_2)$	$\ge 1 - \epsilon$	$\le \epsilon$	$\le \epsilon$
$C' = (h', P_1', P_2')$	$\le \epsilon$	$\ge 1 - \epsilon$	$\le \epsilon$

$$Z_h(n) = \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{i, j \in \mathcal{S}} N_h(n, \underline{d}, \underline{i}, j)\ \log P_1^{d_h}(j \mid i_h) + \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a \neq h} \sum_{i, j \in \mathcal{S}} N_a(n, \underline{d}, \underline{i}, j)\ \log P_2^{d_h}(j \mid i_h)$$

$$+ \log P_C(A_0, B_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t, B_t \mid B_0^{t-1}, A_0^{t-1}, \bar{X}_0^{t-1})$$

SRS Policy

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \le s \le t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

- π is a stationary randomised strategy (SRS policy in short) if $\exists\ \lambda(\,\cdot \mid \cdot\,)$ such that

$$P(B_t \mid B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \le s \le t\}) = \lambda(B_t \mid \underline{d}(t), \underline{i}(t))$$

- Such an SRS policy will be denoted as π^λ
- Π_{SRS} : set of all SRS policies

Ergodicity

- Under an SRS policy π^λ, the process $\{(\underline{d}(t), \underline{i}(t)) : t \geq K\}$ is a Markov process

- Thanks to the trembling hand, the above Markov process is ergodic

- Let $\mu^\lambda = \{\mu^\lambda(\underline{d}, \underline{i}) : (\underline{d}, \underline{i}) \in \mathbb{S}\}$ be the stationary distribution for π^λ

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

ergodic state-action occupancy

$R^*(P_1, P_2)$ in More Detail

$$R^*(P_1,P_2) = \sup_{\nu} \min_{h' \neq h} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d},\underline{i},a)\, k_{hh'}(\underline{d},\underline{i},a)$$

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

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

$$\nu(\underline{d},\underline{i},a) \geq 0 \quad \forall\ (\underline{d},\underline{i},a)$$

Difficult to show that this supremum is attained

Theorem 8.8.2, Puterman[3]

$$R^*(P_1,P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d},\underline{i},a)\, k_{hh'}(\underline{d},\underline{i},a)$$

3. M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, 2014.

δ-Optimal Solutions

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\ k_{hh'}(\underline{d}, \underline{i}, a)$$

- Computability of the sup is an issue.
  Q-learning may be needed.

- For $\delta > 0$, under $C = (h, P_1, P_2)$, let $\lambda_{h,P_1,P_2,\delta}$ be such that

$$\min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^{\lambda_{h,P_1,P_2,\delta}}(\underline{d}, \underline{i}, a)\ k_{hh'}(\underline{d}, \underline{i}, a)) \geq \frac{R^*(P_1, P_2)}{1 + \delta}$$

δ-optimal solution for $C = (h, P_1, P_2)$

Policy $\pi_1^\star(L, \delta)$

$$Z_h(n) = \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{i,j\in\mathcal{S}} N_h(n,\underline{d},\underline{i},j)\ \log P_1^{d_h}(j\,|\,i_h) + \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a\neq h} \sum_{i,j\in\mathcal{S}} N_a(n,\underline{d},\underline{i},j)\ \log P_2^{d_h}(j\,|\,i_h)$$
$$+\log P_C(A_0, B_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t, B_t\,|\,B_0^{t-1}, A_0^{t-1}, \bar{X}_0^{t-1})$$

$$R^*(P_1, P_2) = \sup_{\pi^\lambda\in\Pi_{SRS}} \min_{h'\neq h} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d},\underline{i},a)\ k_{hh'}(\underline{d},\underline{i},a)$$

- Select each arm once ($n = 0,\ldots,K-1$)
- For $n \geq K$, repeat the following until stoppage:
    - Estimate $\hat{h}(n)$
    - If $\min_{h'\neq\hat{h}(n)} Z_{\hat{h}(n),h'}(n) \geq \log((K-1)L)$, stop and declare $\hat{h}(n)$ as the odd arm
    - Else, sample next arm according to $\lambda^*_{\hat{h}(n),\ P_1,\ P_2,\ \delta}(\cdot\,|\,\underline{d}(n), \underline{i}(n))$

$$h \mapsto \lambda^*_{h,P_1,P_2,\delta}$$

$$\hat{h}(n) \in \arg\max_h \min_{h'\neq h} Z_{hh'}(n)$$

Performance of $\pi_1^\star(L,\delta)$

- Stops in finite time w.p. 1
- $\hat{h}(n) = h$ for all n large, almost surely
- If $L = 1/\epsilon$, then $\pi_1^\star(L,\delta) \in \Pi(\epsilon)$
- Upper bound:

$$\limsup_{L\to\infty} \frac{E[\tau(\pi_1^\star(L,\delta)) \mid C]}{\log L} \le \frac{1+\delta}{R^*(P_1,P_2)}$$

- Therefore,

$$\lim_{\delta\downarrow 0} \limsup_{L\to\infty} \frac{E[\tau(\pi_1^\star(L,\delta)) \mid C]}{\log L} \le \frac{1}{R^*(P_1,P_2)}$$

$$C = (h, P_1, P_2)$$

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

$$R^*(P_1,P_2) = \sup_{\pi^\lambda\in\Pi_{SRS}} \min_{h'\neq h} \sum_{(\underline{d},\underline{i})\in\mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d},\underline{i},a)\, k_{hh'}(\underline{d},\underline{i},a)$$

Part 3: Restless Arms with TPMs Unknown

P. N. Karthik and Rajesh Sundaresan, “Learning to Detect an Odd Restless Markov Arm with a Trembling Hand”, submitted.

Learning to Detect an Odd Restless Markov Arm

- A multi-armed bandit with $K \geq 3$ arms
- Each arm is a time homogeneous and ergodic Markov process
- Markov processes evolve on a common, finite state space
- The TPM of one of the arms (odd arm) is P_1; TPM of rest of the arms is P_2
- Arms are restless
- TPMs are unknown (learning)

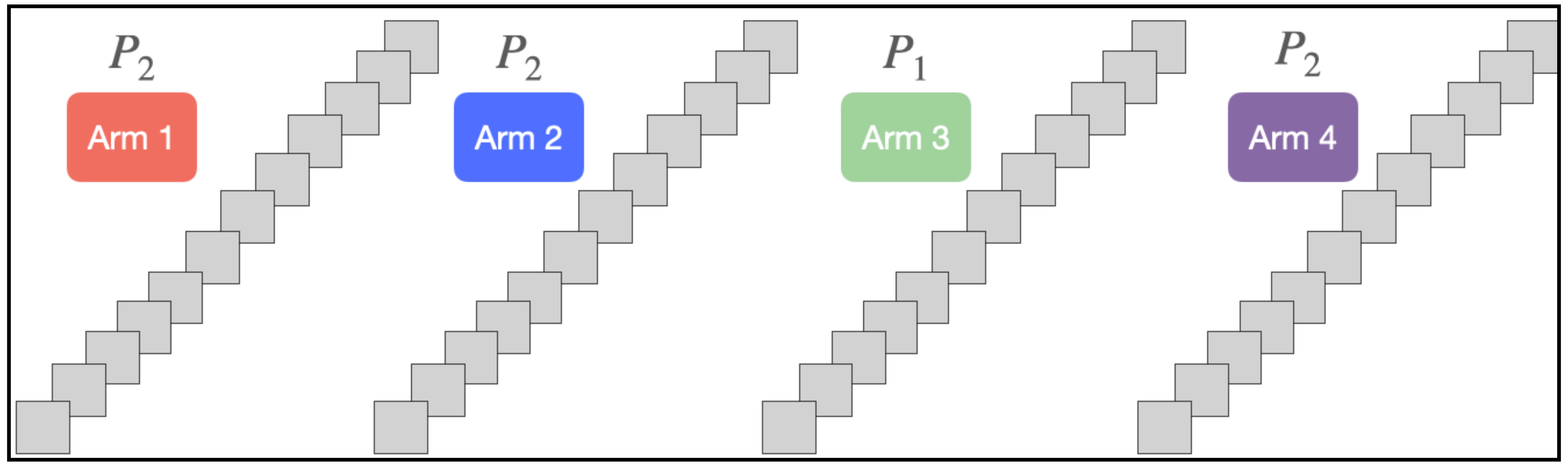
Characterise $$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)}$$

Our Contributions

- Let $C = (h, P_1, P_2)$ be a problem instance
- Lower bound:

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

- Policy — matching upper bound as $\epsilon \downarrow 0$ under
  - Continuous selection assumption
  - Regularity assumption on the TPMs



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

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{C':\, h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

Markov Decision Problem (MDP)

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \le s \le t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

MDP Transition Probabilities

$$\underline{d}(t) = \underline{d} = (4,3,2,1) \qquad \underline{i}(t) = \underline{i} = (i_1, i_2, i_3, i_4)$$

$$B_t = b \qquad A_t = 1$$

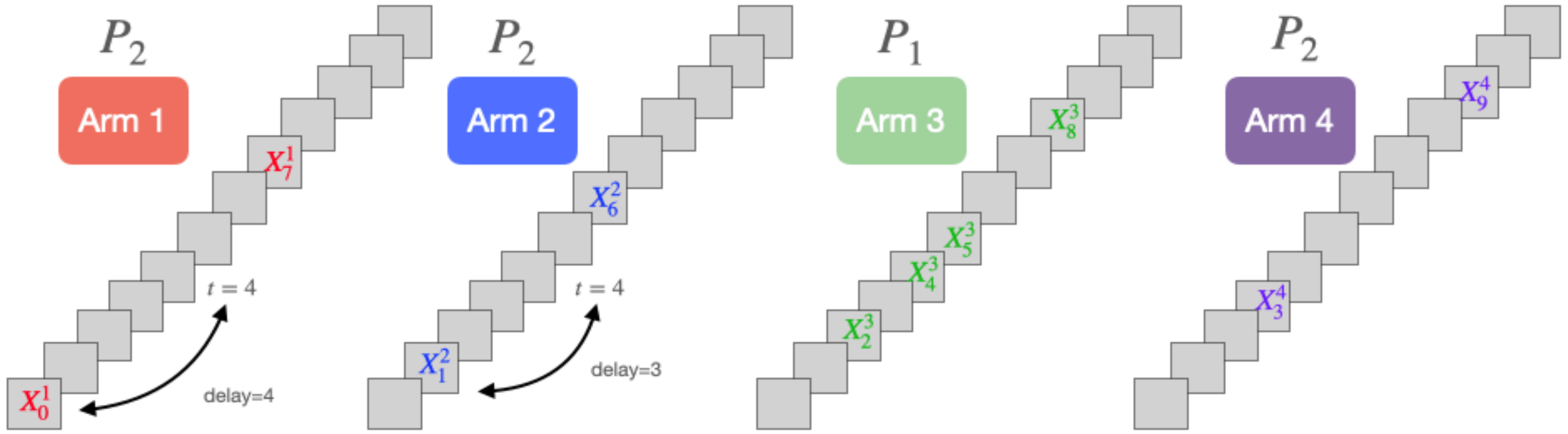
$$P(A_t = 1 \,|\, B_t = b) = \frac{\eta}{K} + (1-\eta)\,\mathbb{1}_{\{b=1\}}$$

$$\underline{d}(t+1) = \underline{d}' = (1,4,3,2) \qquad \underline{i}(t+1) = \underline{i}' = (X_t^1, i_2, i_3, i_4)$$

$$P(\underline{d}(t+1) = \underline{d}', \underline{i}(t+1) = \underline{i}' \,|\, \underline{d}(t) = \underline{d}, \underline{i}(t) = \underline{i}, B_t = b) = \left(\frac{\eta}{K} + (1-\eta)\,\mathbb{1}_{\{b=1\}}\right)(P_2)^4(X_t^1 \,|\, i_1)$$

$$Q(\underline{d}'', \underline{i}'' \,|\, \underline{d}', \underline{i}', b)$$

Fourth power of the TPM P_2



t	A_t	Arm 1		Arm 2		Arm 3		Arm 4	
		$d_1(t)$	$i_1(t)$	$d_2(t)$	$i_2(t)$	$d_3(t)$	$i_3(t)$	$d_4(t)$	$i_4(t)$
0	1								
1	2								
2	3								
3	4								
4	3	4	X_0^1	3	X_1^2	2	X_2^3	1	X_3^4
5	3	5	X_0^1	4	X_1^2	1	X_4^3	2	X_3^4
6	2	6	X_0^1	5	X_1^2	1	X_5^3	3	X_3^4
7	1	7	X_0^1	1	X_6^2	2	X_5^3	4	X_3^4
8	3	1	X_7^1	2	X_6^2	3	X_5^3	5	X_3^4
9		2	X_7^1	3	X_6^2	1	X_8^3	6	X_3^4

X_t^a : observation from arm a at time t $\quad$ $d_a(t)$: delay of arm a at time t $\quad$ $i_a(t)$: last observed state of arm a at time t

Characterise $\displaystyle \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)}$

MDP Transition Probabilities

- The MDP transition probabilities are parameterised by the arms configuration
- The value of the true parameter (underlying arms configuration) is unknown and must be learnt (identification / identifiability)
- The set of all possible parameters is uncountably infinite

Data processing inequality

$$d(\epsilon, 1-\epsilon) \leq E[Z_C(\tau(\pi)) - Z_{C'}(\tau(\pi)) \mid C] \lesssim E[\tau(\pi) \mid C] \cdot R^*(P_1, P_2)$$

Information theoretic bottleneck: maximum discrimination per unit time

$C = (h, P_1, P_2)$

$C' = (h', P_1', P_2')$ $\quad h' \neq h$

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

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{C':\, h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

$$k_{CC'}(\underline{d}, \underline{i}, a) = \begin{cases} D(P_1^{d_a}(\cdot \mid i_a) \| (P_2')^{d_a}(\cdot \mid i_a)), & a = h, \\ D(P_2^{d_a}(\cdot \mid i_a) \| (P_1')^{d_a}(\cdot \mid i_a)), & a = h', \\ D(P_2^{d_a}(\cdot \mid i_a) \| (P_2')^{d_a}(\cdot \mid i_a)), & a \neq h, h', \end{cases}$$

$$\pi \in \Pi(\epsilon)$$

Configuration	Decision $= h$	Decision $= h'$	Others
$C = (h, P_1, P_2)$	$\geq 1 - \epsilon$	$\leq \epsilon$	$\leq \epsilon$
$C' = (h', P_1', P_2')$	$\leq \epsilon$	$\geq 1 - \epsilon$	$\leq \epsilon$

$$Z_C(n) = \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{i,j \in \mathcal{S}} N_h(n, \underline{d}, \underline{i}, j)\ \log P_1^{d_h}(j \mid i_h) + \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n, \underline{d}, \underline{i}, j)\ \log P_2^{d_h}(j \mid i_h)$$

$$+ \log P_C(A_0, B_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^{n} \log P_C(A_t, B_t \mid B_0^{t-1}, A_0^{t-1}, \bar{X}_0^{t-1})$$

δ-Optimal Solutions

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{C':\, h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

- Computability of the sup is an issue.
  Q-learning may be needed.

- For $\delta > 0$, under $C = (h, P_1, P_2)$, let $\lambda_{h,P_1,P_2,\delta}$ be such that

$$\min_{C':\, h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^{\lambda_{h,P_1,P_2,\delta}}(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)) \geq \frac{R^*(P_1, P_2)}{1 + \delta}$$

δ-optimal solution for $C = (h, P_1, P_2)$

δ-Optimal Solutions

Fix $\delta > 0$

(h''', P''', Q''')

$\vdots$

(h'', P'', Q'')

$\vdots$

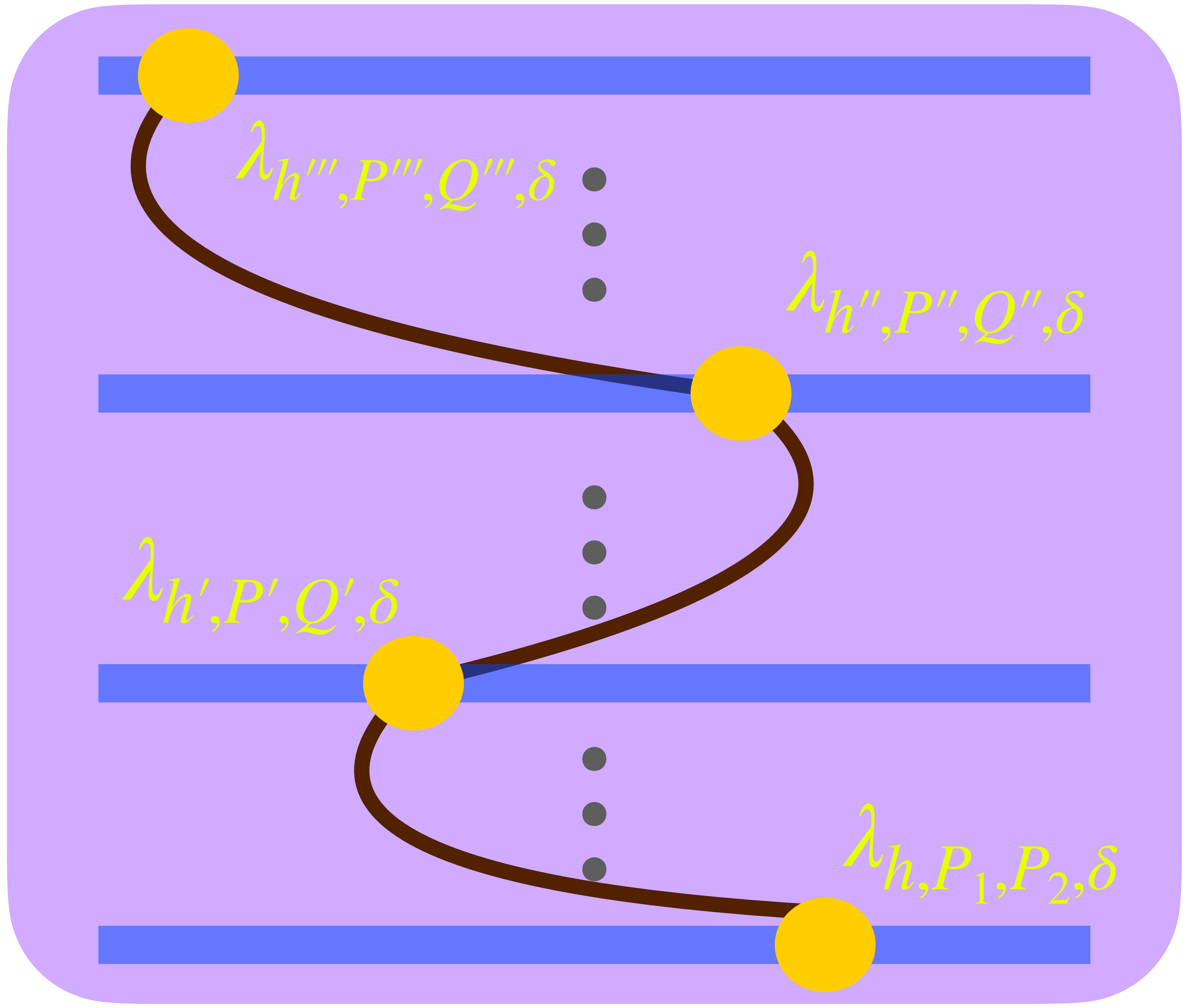
(h', P', Q')

$\vdots$

(h, P_1, P_2)

Set of all possible arms configurations (parameters)

$$(h, P, Q) \mapsto \lambda_{h,P,Q,\delta}$$



$\{\lambda_{h,P,Q,\delta}\}_{h,P,Q}$

$$(\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),2}(n)) \approx (h, P_1, P_2)$$

$$\lambda_{\hat{h}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \hat{P}_{\hat{h}(n),1}(n),\ \delta} \approx \lambda_{h,P_1,P_2,\delta}$$

Two Key Assumptions

For each $\delta > 0$, there exists a selection $\{\lambda_{h,P,Q,\delta}\}_{h,P,Q}$ such that $(h, P, Q) \mapsto \lambda_{h,P,Q,\delta}$ is continuous

$$\mathscr{P}(\bar{\varepsilon}^*) = \{P : \ P \text{ is ergodic, } P^d(j\,|\,i) > 0 \implies P^d(j\,|\,i) \geq \bar{\varepsilon}^* \text{ for all } d \geq 1,\ i, j\}$$

There exists $\bar{\varepsilon}^* \in (0,1)$ such that for any (h, P, Q), the TPMs $P, Q \in \mathscr{P}(\bar{\varepsilon}^*)$

$P, Q \in \mathscr{P}(\bar{\varepsilon}^*)$ are harder to distinguish than otherwise

Arbitrary P, Q

$$0 \leq D(P^d(\,\cdot\,|\,i) \| Q^d(\,\cdot\,|\,i)) \leq \infty$$

$P, Q \in \mathscr{P}(\bar{\varepsilon}^*)$

$$0 \leq D(P^d(\,\cdot\,|\,i) \| Q^d(\,\cdot\,|\,i)) \leq \log \frac{1}{\bar{\varepsilon}^*}$$

Policy $\pi_2^\star(L, \delta)$

- For $n = 0, \ldots, K-1$, sample each of the K arms once
- For all $n \geq K$, repeat the following steps until stoppage:
  - Compute ML estimates $(\hat{P}_1(n), \hat{P}_2(n))$ of the TPMs [no closed-form expressions]
  - Let

$$\hat{h}(n) \in \arg\max_{h} \underbrace{\min_{h' \neq h} \log \frac{\text{avg. likelihood up to time } n \text{ when } h \text{ is the odd arm}}{\text{max. likelihood up to time } n \text{ when } h' \text{ is the odd arm}}}_{M_h(n)}$$

  - If $M_{\hat{h}(n)}(n) \geq \log((K-1)L)$, stop and declare $\hat{h}(n)$ as the odd arm
  - Else, sample the next arm according to $\lambda_{\hat{h}(n), \hat{P}_1(n), \hat{P}_2(n), \delta}(\,\cdot \mid \underline{d}(n), \underline{i}(n))$
  - Update $n \leftarrow n+1$

Principle of certainty equivalence

Performance of $\pi_2^\star(L, \delta)$

For each $\delta > 0$, there exists a selection $\{\lambda_{h,P,Q,\delta}\}_{h,P,Q}$ such that $(h, P, Q) \mapsto \lambda_{h,P,Q,\delta}$ is continuous

There exists $\bar{\varepsilon}^* \in (0,1)$ such that for any $C = (h, P, Q)$, the TPMs $P, Q \in \mathscr{P}(\bar{\varepsilon}^*)$

- $(\hat{h}(n), \hat{P}_1(n), \hat{P}_2(n)) \to (h, P_1, P_2)$ (identification/identifiability)
- If $L = 1/\epsilon$, then $\pi_2^\star(L, \delta) \in \Pi(\epsilon)$ for all $\delta > 0$
- Under $C = (h, P_1, P_2)$, we have

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{C'} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\, k_{CC'}(\underline{d}, \underline{i}, a)$$

$$\limsup_{L \to \infty} \frac{E[\tau(\pi_2^\star(L, \delta)) \,|\, C]}{\log L} \le \frac{(1+\delta)^2}{R^*(P_1, P_2)}$$

$$\lim_{\delta \downarrow 0} \limsup_{L \to \infty} \frac{E[\tau(\pi_2^\star(L, \delta)) \,|\, C]}{\log L} \le \frac{1}{R^*(P_1, P_2)}$$

In a Nutshell

Rested Arms, Unknown TPMs

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)} = \frac{1}{D^*(h, P_1, P_2)}$$

Restless Arms, Known TPMs

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

Restless Arms, Unknown TPMs

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

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

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\ k_{hh'}(\underline{d}, \underline{i}, a)$$

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{C':\, h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\ k_{CC'}(\underline{d}, \underline{i}, a)$$

$$\lim_{n \to \infty} \frac{\#\text{ transitions from } i}{n} = \lim_{n \to \infty} \frac{\#\text{ transitions to } i}{n}$$

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

There exists $\bar{\varepsilon}^* \in (0,1)$ such that for any (h, P, Q), the TPMs $P, Q \in \mathscr{P}(\bar{\varepsilon}^*)$

For each $\delta > 0$, there exists a selection $\{\lambda_{h,P,Q,\delta}\}_{h,P,Q}$ such that $(h, P, Q) \mapsto \lambda_{h,P,Q,\delta}$ is continuous

Future Work

The Case $\eta = 0$

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \mu^\lambda(\underline{d}, \underline{i}) \cdot \left(\frac{\eta}{K} + (1-\eta)\ \lambda(a \mid \underline{d}, \underline{i}) \right) k_{hh'}(\underline{d}, \underline{i}, a)$$

$$\frac{1}{R^*(P_1, P_2)} \leq \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} \leq \lim_{\delta \downarrow 0} \limsup_{L \to \infty} \frac{E[\tau(\pi_1^\star(L, \delta)) \mid C]}{\log L} \leq \frac{1}{R^*(P_1, P_2)}$$

$$\frac{1}{R_\eta^*(P_1, P_2)} \leq \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} \leq \lim_{\delta \downarrow 0} \limsup_{L \to \infty} \frac{E[\tau(\pi_1^\star(L, \delta)) \mid C]}{\log L} \leq \frac{1}{R_\eta^*(P_1, P_2)}$$

What happens as $\eta \downarrow 0$?

The Case $\eta = 0$

$$R^*_\eta(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \mu^\lambda(\underline{d}, \underline{i}) \cdot \left(\frac{\eta}{K} + (1-\eta)\, \lambda(a \,|\, \underline{d}, \underline{i}) \right) \cdot k_{hh'}(\underline{d}, \underline{i}, a)$$

Monotonicity: $\eta' < \eta \implies R^*_\eta(P_1, P_2) \leq R^*_{\eta'}(P_1, P_2)$

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

$$R^*_0(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \mu^\lambda(\underline{d}, \underline{i}) \cdot \lambda(a \,|\, \underline{d}, \underline{i}) \cdot k_{hh'}(\underline{d}, \underline{i}, a)$$

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

Envelope theorem

IID observations from the arms ✓

Rested arms ✓

Restless arms : $\lim_{\eta \downarrow 0} R^*_\eta(P_1, P_2) \leq R^*_0(P_1, P_2)$

The Case $\eta = 0$

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \mu^\lambda(\underline{d}, \underline{i}) \cdot \left(\frac{\eta}{K} + (1-\eta)\, \lambda(a \,|\, \underline{d}, \underline{i}) \right) k_{hh'}(\underline{d}, \underline{i}, a)$$

$$\frac{1}{R^*(P_1, P_2)} \leq \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)} \leq \lim_{\delta \downarrow 0} \limsup_{L \to \infty} \frac{E[\tau(\pi_1^\star(L, \delta)) \,|\, C]}{\log L} \leq \frac{1}{R^*(P_1, P_2)}$$

$$\frac{1}{R_\eta^*(P_1, P_2)} \leq \lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)} \leq \lim_{\delta \downarrow 0} \limsup_{L \to \infty} \frac{E[\tau(\pi_1^\star(L, \delta)) \,|\, C]}{\log L} \leq \frac{1}{R_\eta^*(P_1, P_2)}$$

What happens as $\eta \downarrow 0$?

$$\frac{1}{R_0^*(P_1, P_2)}$$

$$\frac{1}{\lim_{\eta \downarrow 0} R_\eta^*(P_1, P_2)}$$

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

Computability of $R^*(P_1, P_2)$

$$R^*(P_1, P_2) = \sup_{\lambda(\cdot|\cdot)} \min_{h' \neq h} \sum_{(\underline{d},\underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \mu^{\lambda}(\underline{d}, \underline{i}) \cdot \left(\frac{\eta}{K} + (1-\eta)\, \lambda(a \,|\, \underline{d}, \underline{i}) \right) \cdot k_{hh'}(\underline{d}, \underline{i}, a)$$

- Computability of supremum in the above expression is an issue
  $d \in \{1,2,\ldots\}$
- Q-learning for restless bandits[4]
- In practice we may want to impose $d \leq M$ for some large M
  - Forcefully sample an arm if its delay exceeds M
  - How to prove ergodicity?

4. K. Avrachenkov and V. S. Borkar, "Whittle Index Based Q-Learning for Restless Bandits with Average Reward," 2020.

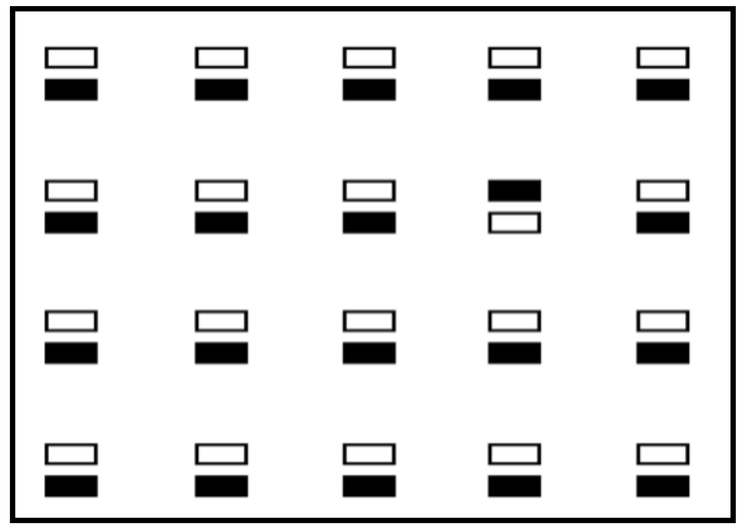
Second Order Term

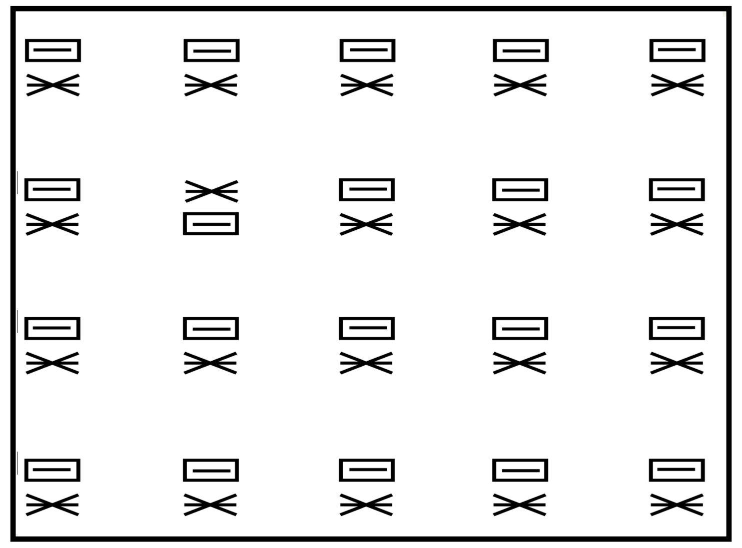
$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, P_1, P_2]}{\log(1/\epsilon)} = \alpha(P_1, P_2)$$

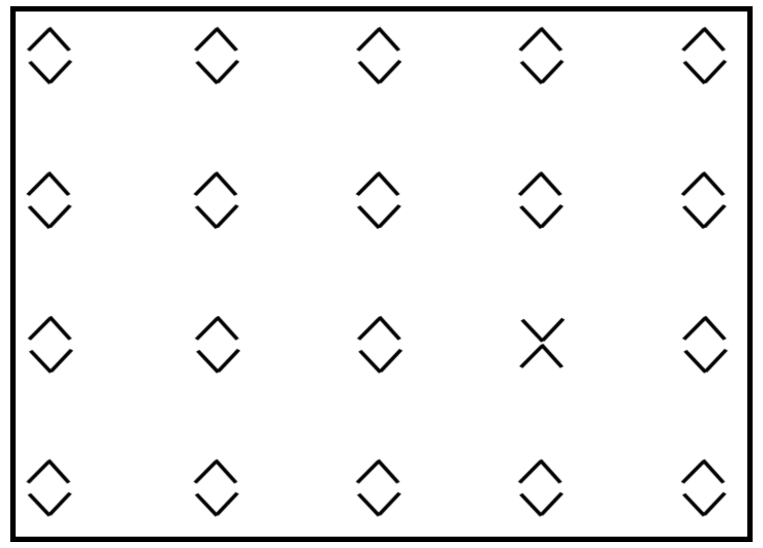
$$\inf_{\pi \in \Pi(\epsilon)} E^{\pi}[\tau(\pi) \,|\, P_1, P_2] \approx \alpha(P_1, P_2) \cdot \log(1/\epsilon)$$

Is there g such that $E^{\pi}[\tau(\pi) \,|\, C] \approx \alpha \cdot \log(1/\epsilon) + \beta \cdot g(\epsilon) + o(g(\epsilon))$?

Future Work (contd.)







- Switching costs
- Sophisticated visual search models

> Grasping multiple objects at once
>
> Memory Constrained Search

- General sequential hypothesis testing (L-anomalous arms identification, best arm identification)

> - Prabhu, Bhashyam, Gopalan, Sundaresan

- Hidden Markov models

G. R. Prabhu, S. Bhashyam, A. Gopalan, and R. Sundaresan, “Sequential Multi- Hypothesis Testing in Multi-Armed Bandit Problems: An Approach for Asymptotic Optimality,” arXiv preprint arXiv:2007.12961, 2020

Acknowledgements

- Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India. (Grant no: EMR/2016/002503)
- Ministry of Human Resource Development, Govt. of India
- Centre for Networked Intelligence, Indian Institute of Science
- Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science

My Heartfelt Thanks

Rajesh Sundaresan
IISc

Srikrishna Bhashyam
IIT Madras

Aditya Gopalan
IISc

Sandeep Juneja
TIFR Mumbai

Navin Kashyap
IISc

Himanshu Tyagi
IISc

Student Brigade

Sarath A Y
PhD student, IISc

Prathamesh Mayekar
PhD student, IISc

Sahasranand K R
PhD student, IISc

Shubhada Agrawal
PhD student, TIFR Mumbai

Gayathri Prabhu
PhD student, IIT Madras

Lakshmi Priya M E
PhD student, IISc

… and many others whose names are not here

Takeaway

Look for invariant quantities

Rested Arms, Unknown TPMs

$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) \,|\, C]}{\log(1/\epsilon)} = \frac{1}{D^*(h, P_1, P_2)}$$

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

$$\lim_{n \to \infty} \frac{\#\text{ transitions from } i}{n} = \lim_{n \to \infty} \frac{\#\text{ transitions to } i}{n}$$

Solve a simpler model
and lift the ideas

Restless Arms, Known TPMs

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

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\ k_{hh'}(\underline{d}, \underline{i}, a)$$

$$(B_0, \ldots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) \to B_t \to (\underline{d}(t+1), \underline{i}(t+1))$$

Identifiability

Search for the right keyword

Restless Arms, Unknown TPMs

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

$$R^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{C':\, h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^{K} \nu^\lambda(\underline{d}, \underline{i}, a)\ k_{CC'}(\underline{d}, \underline{i}, a)$$

There exists $\bar{\varepsilon}^* \in (0,1)$ such that for any (h, P, Q), the TPMs $P, Q \in \mathscr{P}(\bar{\varepsilon}^*)$

For each $\delta > 0$, there exists a selection $\{\lambda_{h,P,Q,\delta}\}_{h,P,Q}$ such that $(h, P, Q) \mapsto \lambda_{h,P,Q,\delta}$ is continuous

Thank You!