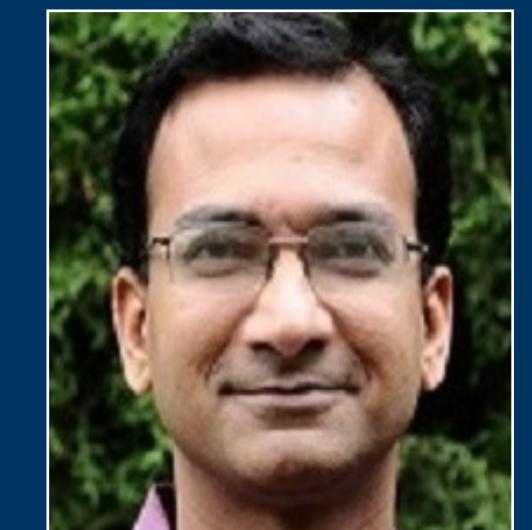


Sequential Controlled Sensing to Detect an Anomalous Process

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



Candidate:
 P. N. Karthik
<https://www.karthikpn.com>



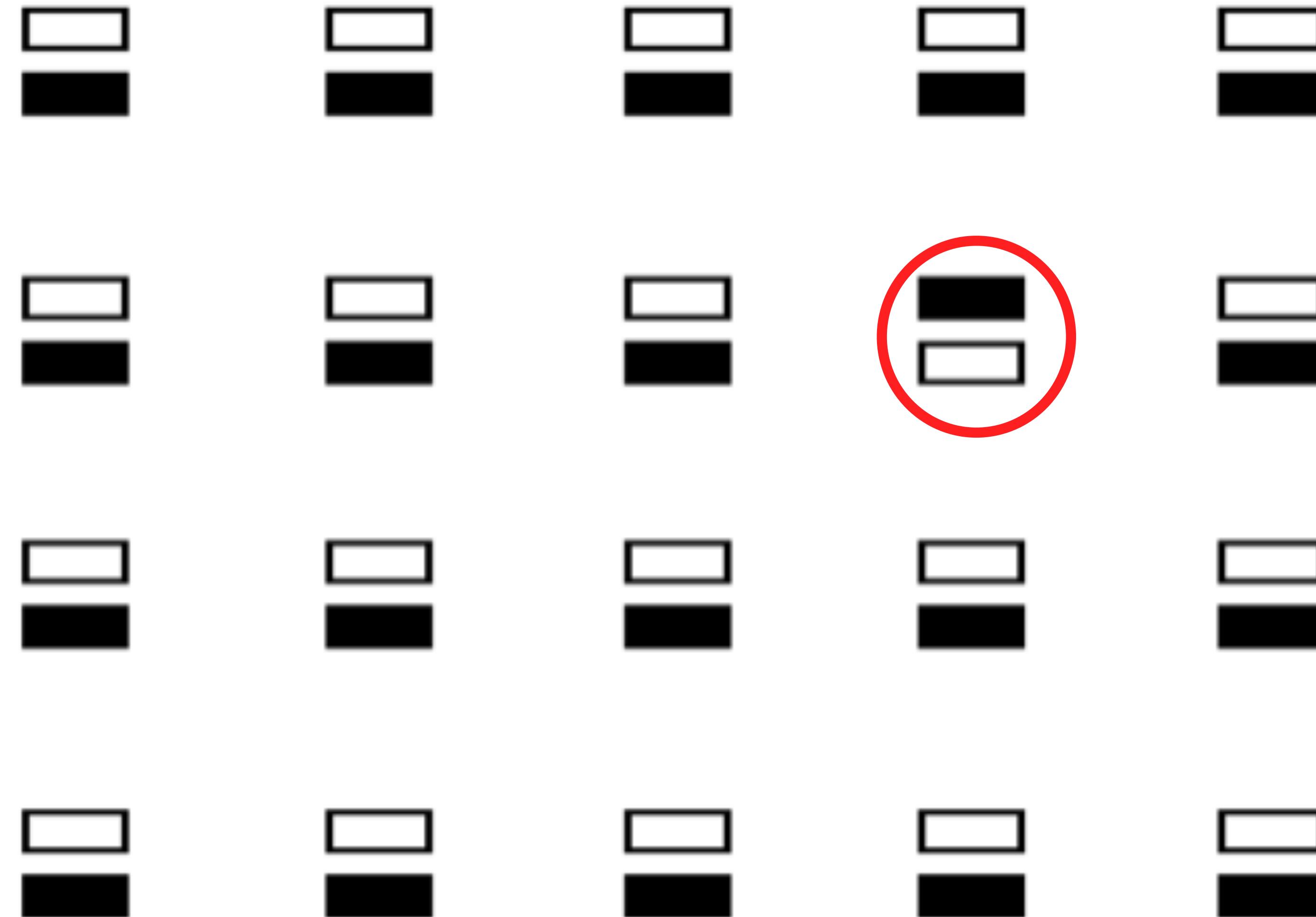
Supervisor:
 Prof. Rajesh Sundaresan
www.ece.iisc.ac.in/~rajeshs

08 November 2021

Motivation

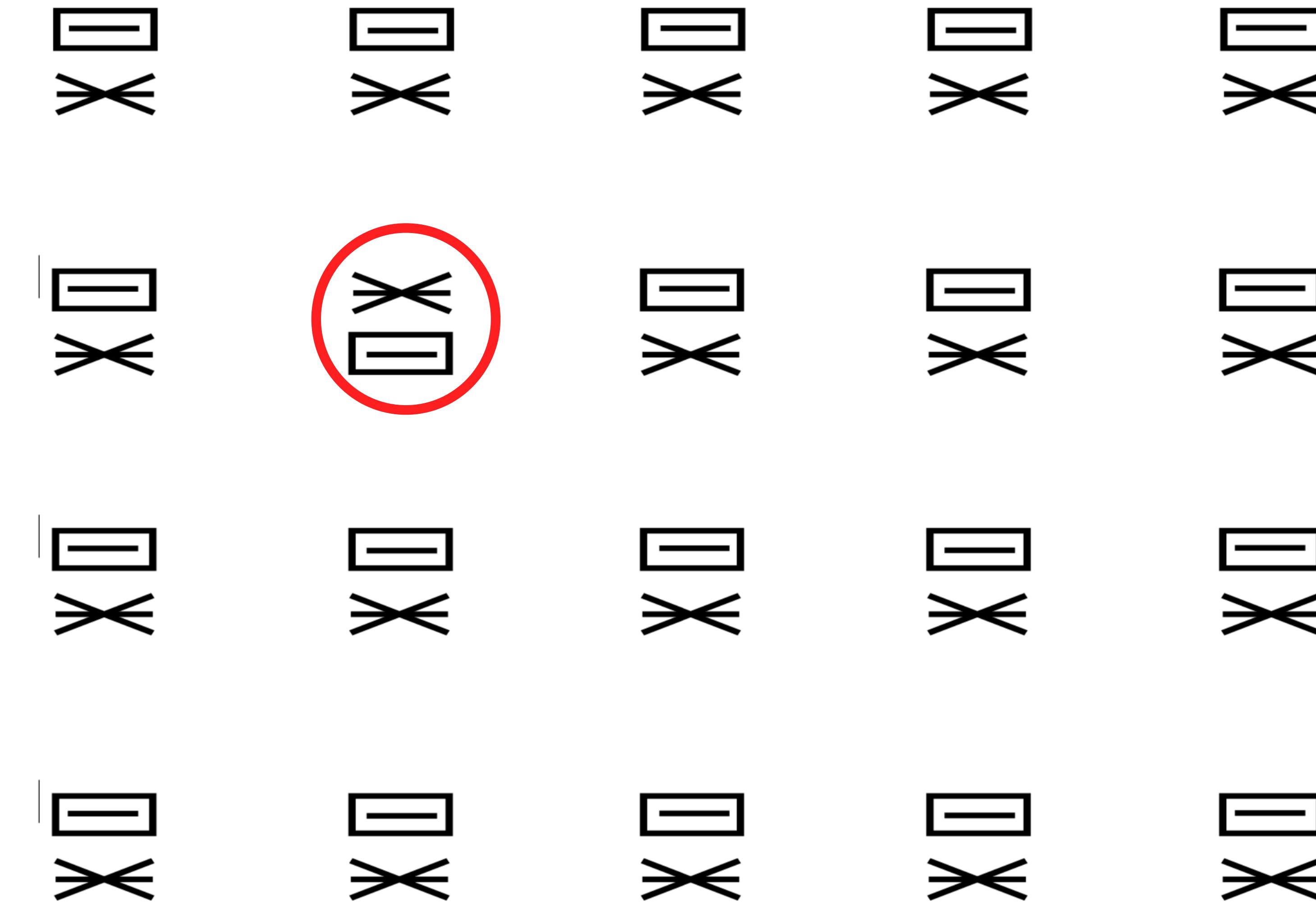
Visual Search Experiments, Multi-Armed Bandits

Find the “Odd” Image – 1



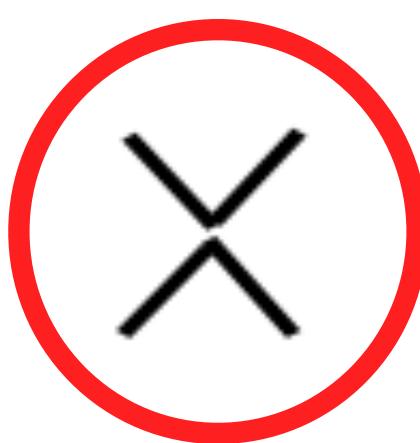
Odd image

Find the “Odd” Image – 2



Odd image

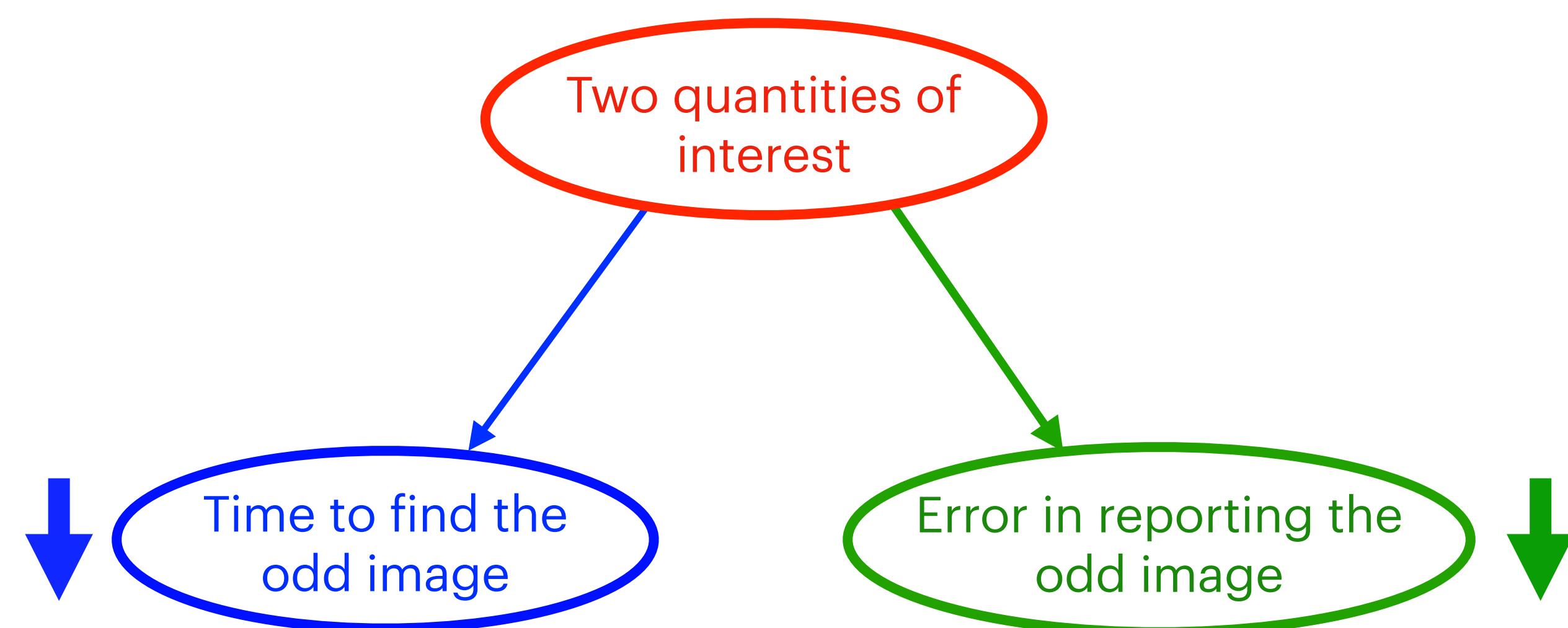
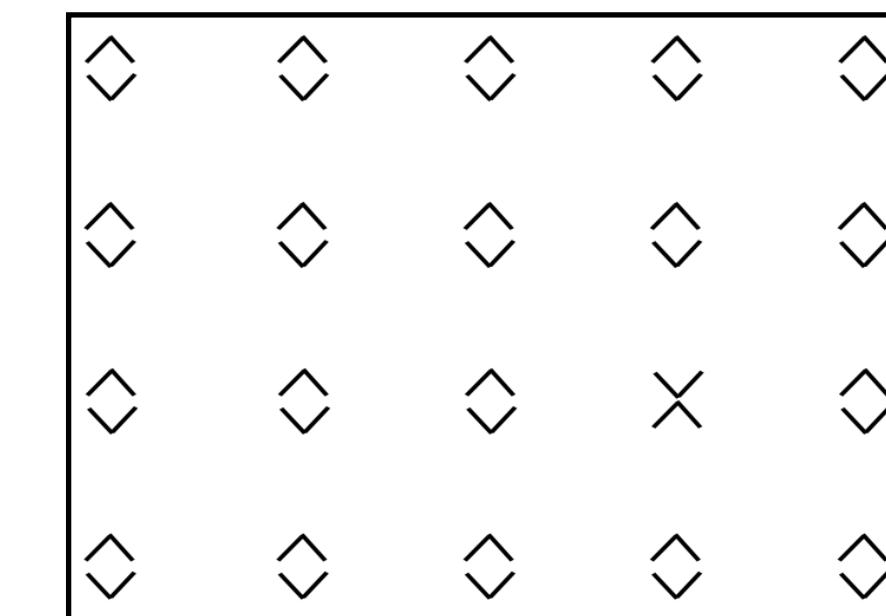
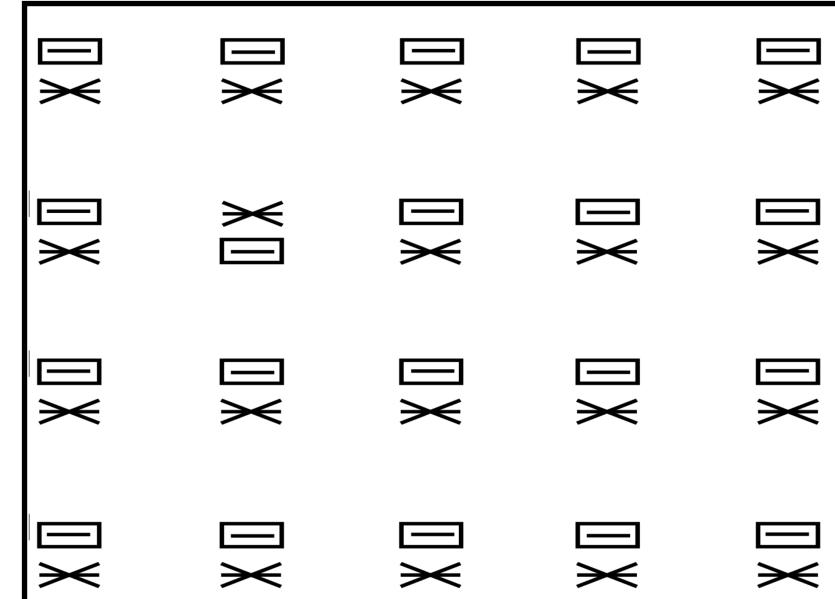
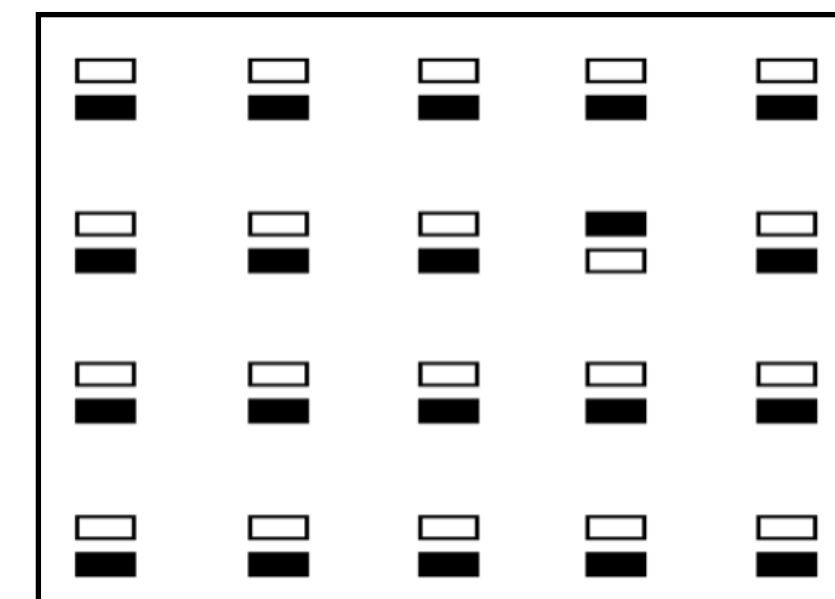
Find the “Odd” Image – 3



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

Goal: to find the odd image quickly and accurately

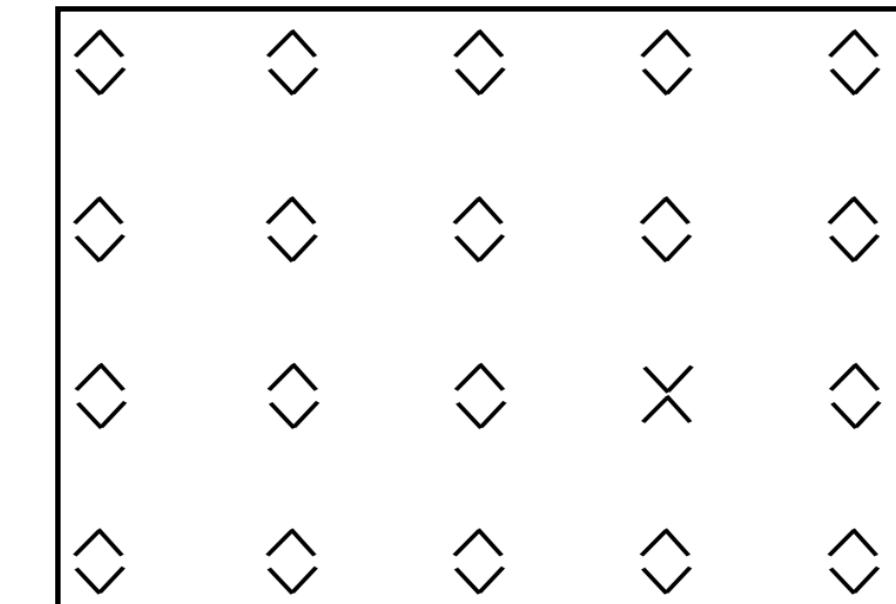
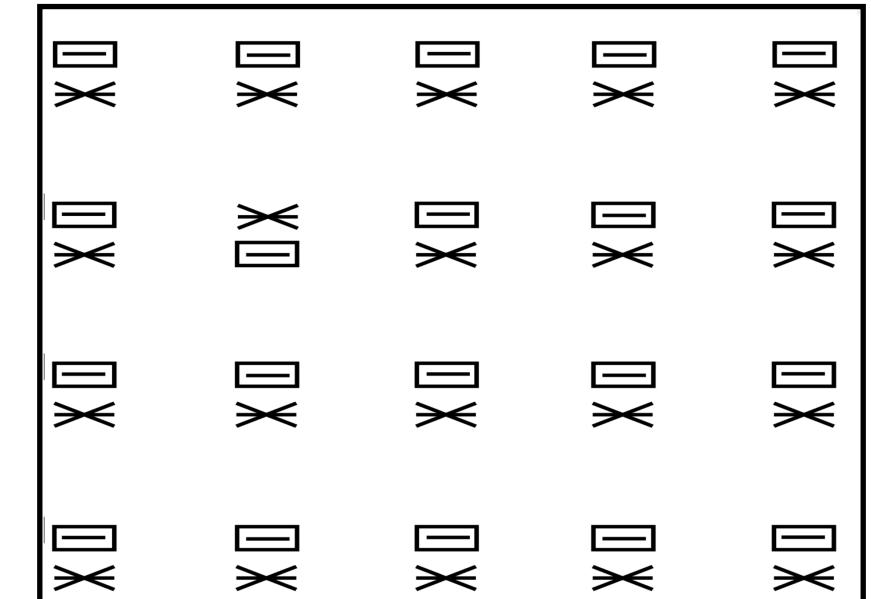
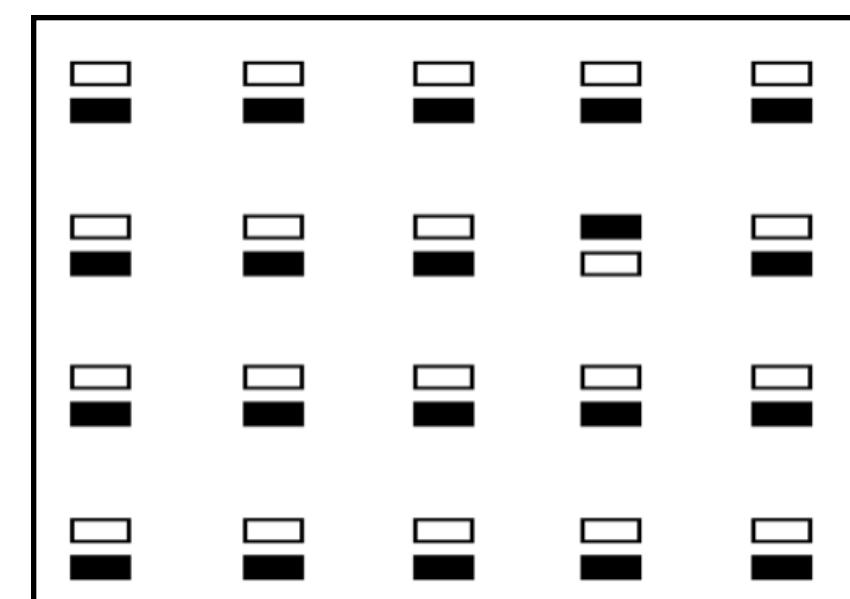
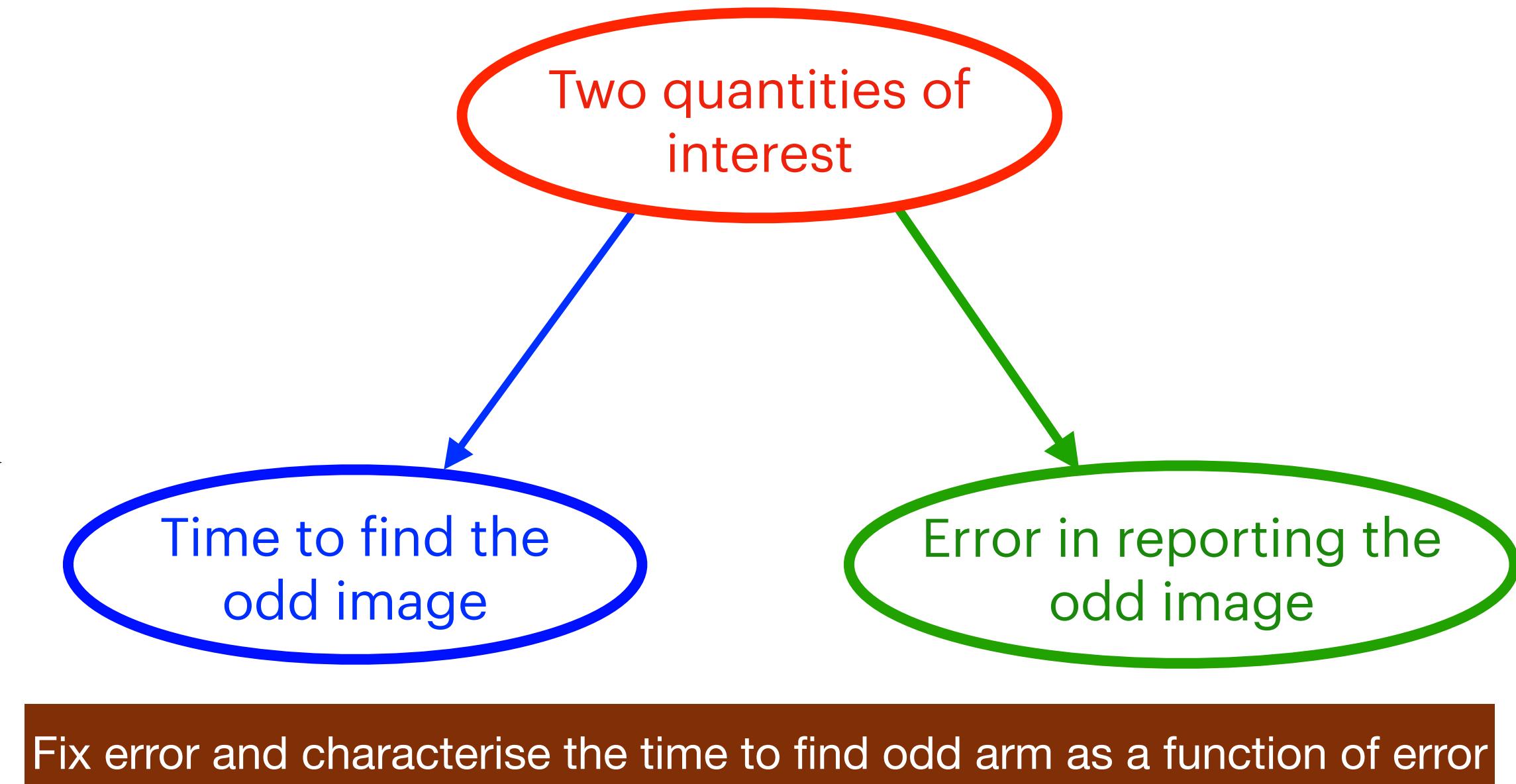
- π : strategy to find the odd image
- $\tau(\pi)$: time to find the odd image under π
- For $\epsilon \in (0,1)$, let $\Pi(\epsilon) = \{\pi : P_{\text{error}}(\pi) \leq \epsilon\}$
- Vaidhiyan et al.^{1,2} showed that for any two image pairs I_1 and I_2 ,

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

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

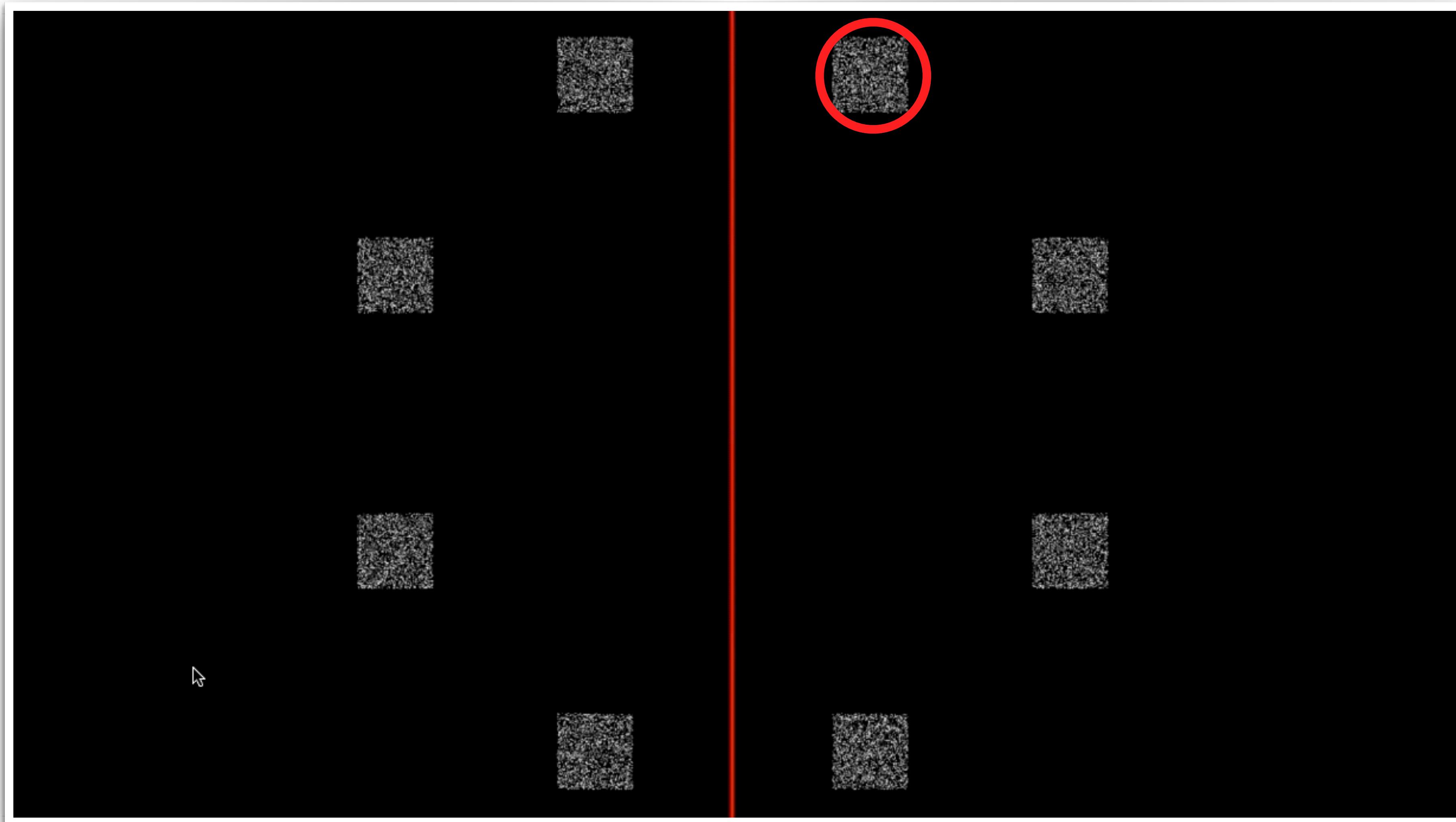
1. Vaidhiyan, N. K., Arun, S. P., & Sundaresan, R. (2017). Neural Dissimilarity Indices that Predict Oddball Detection in Behaviour. *IEEE Transactions on Information Theory*, 63(8), 4778-4796.

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



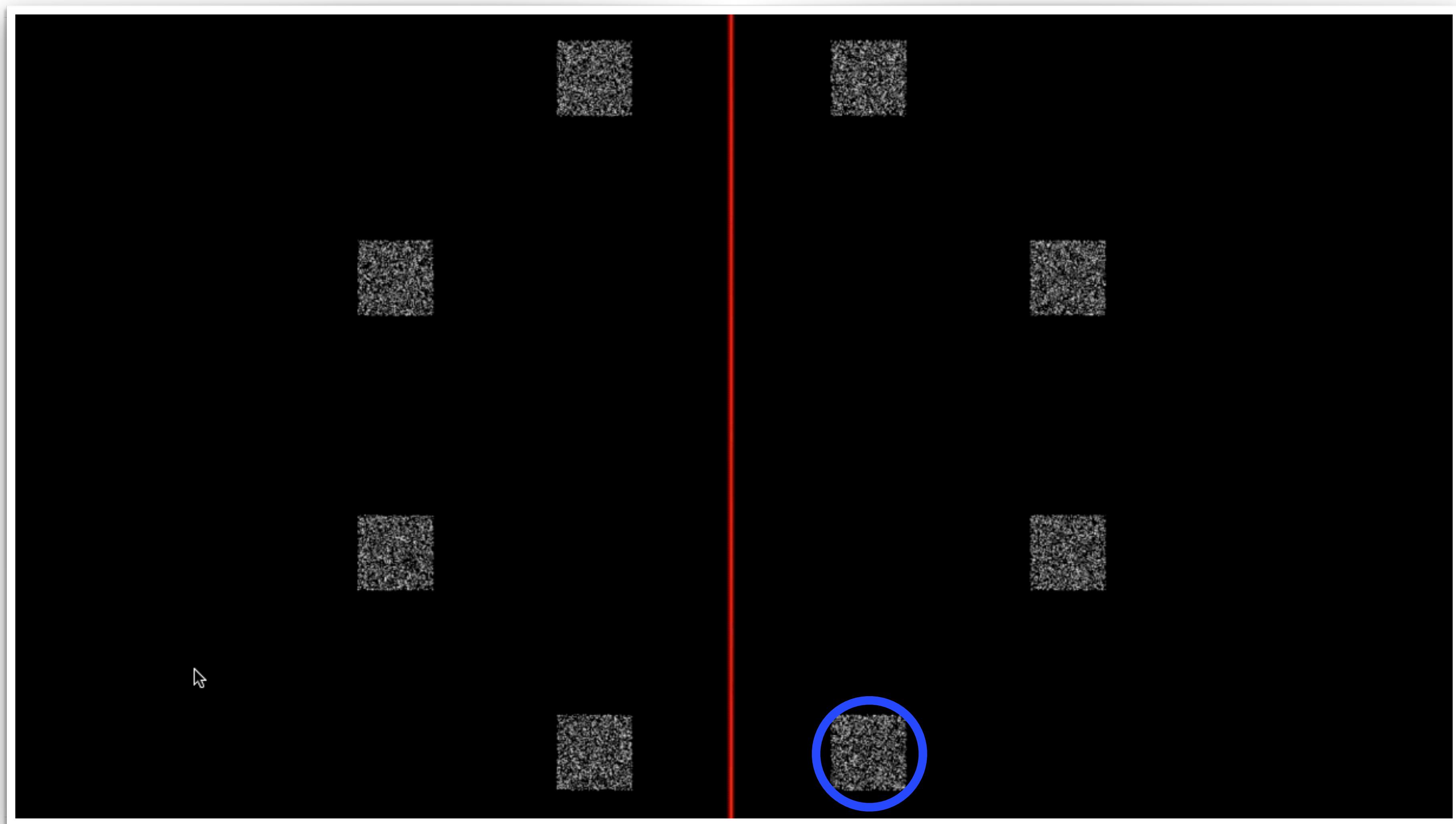
From Static Images to Movies

Find the Odd Movie – 1



Odd movie

Find the Odd Movie – 2



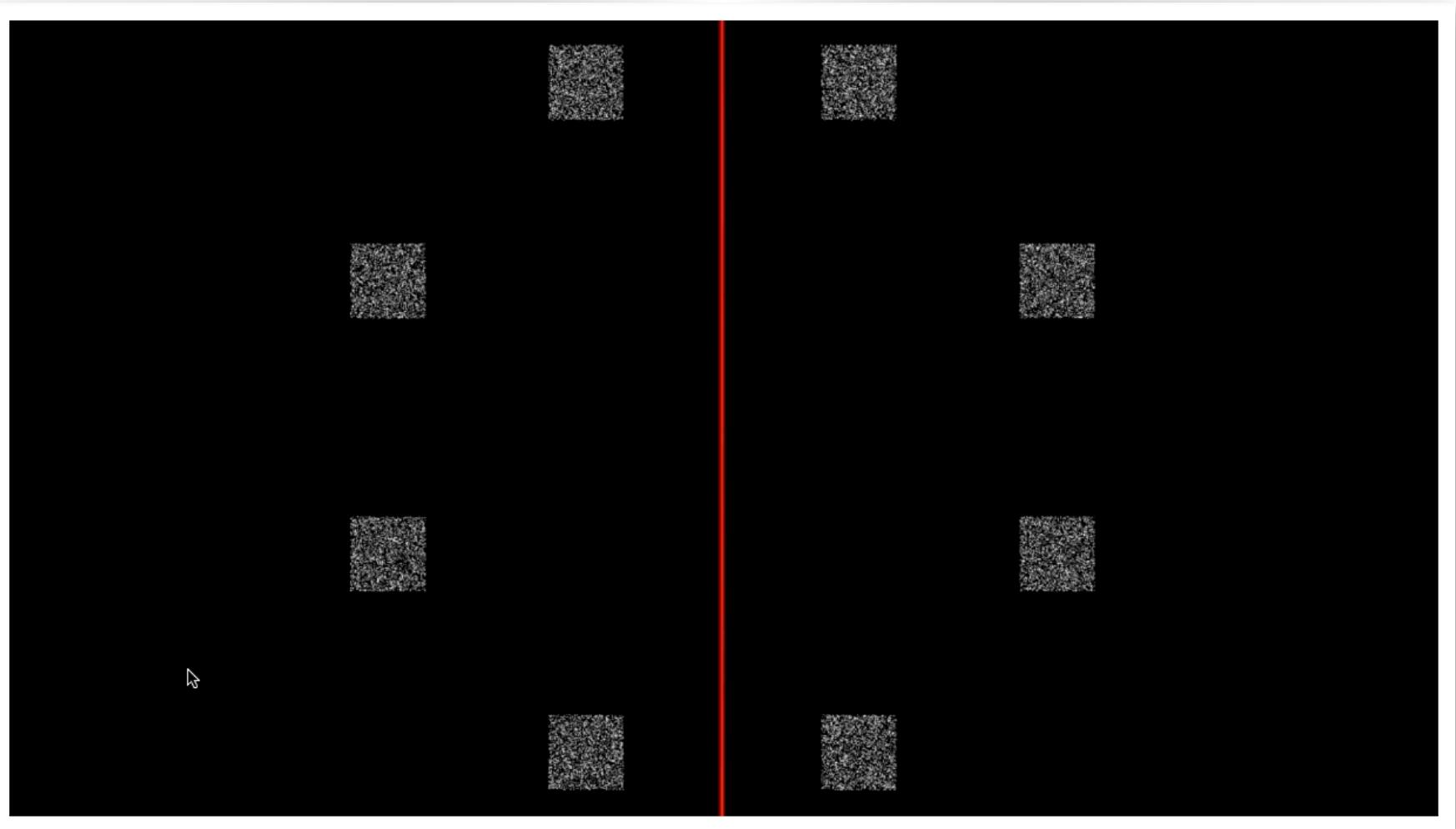
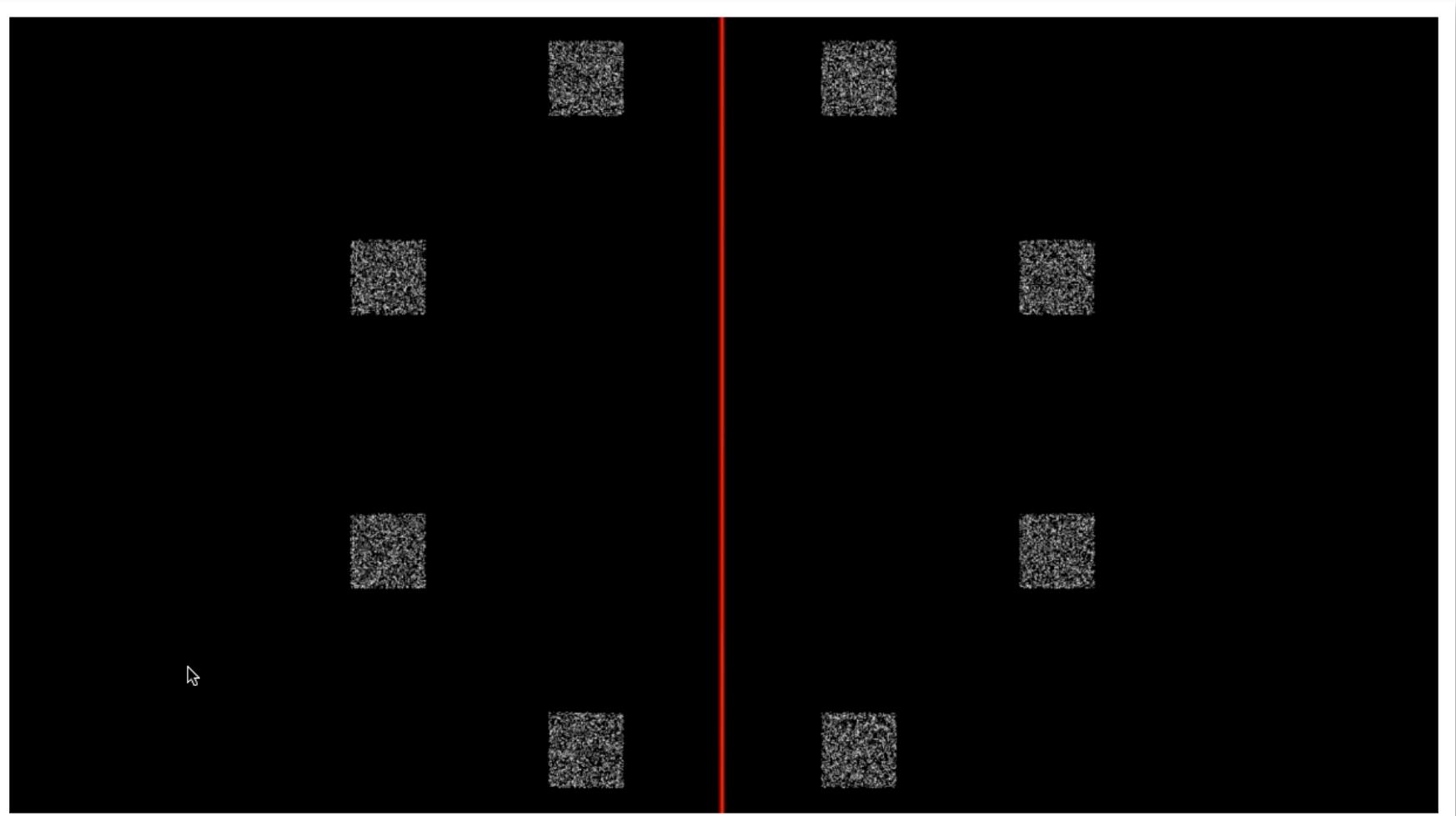
Odd movie

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

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

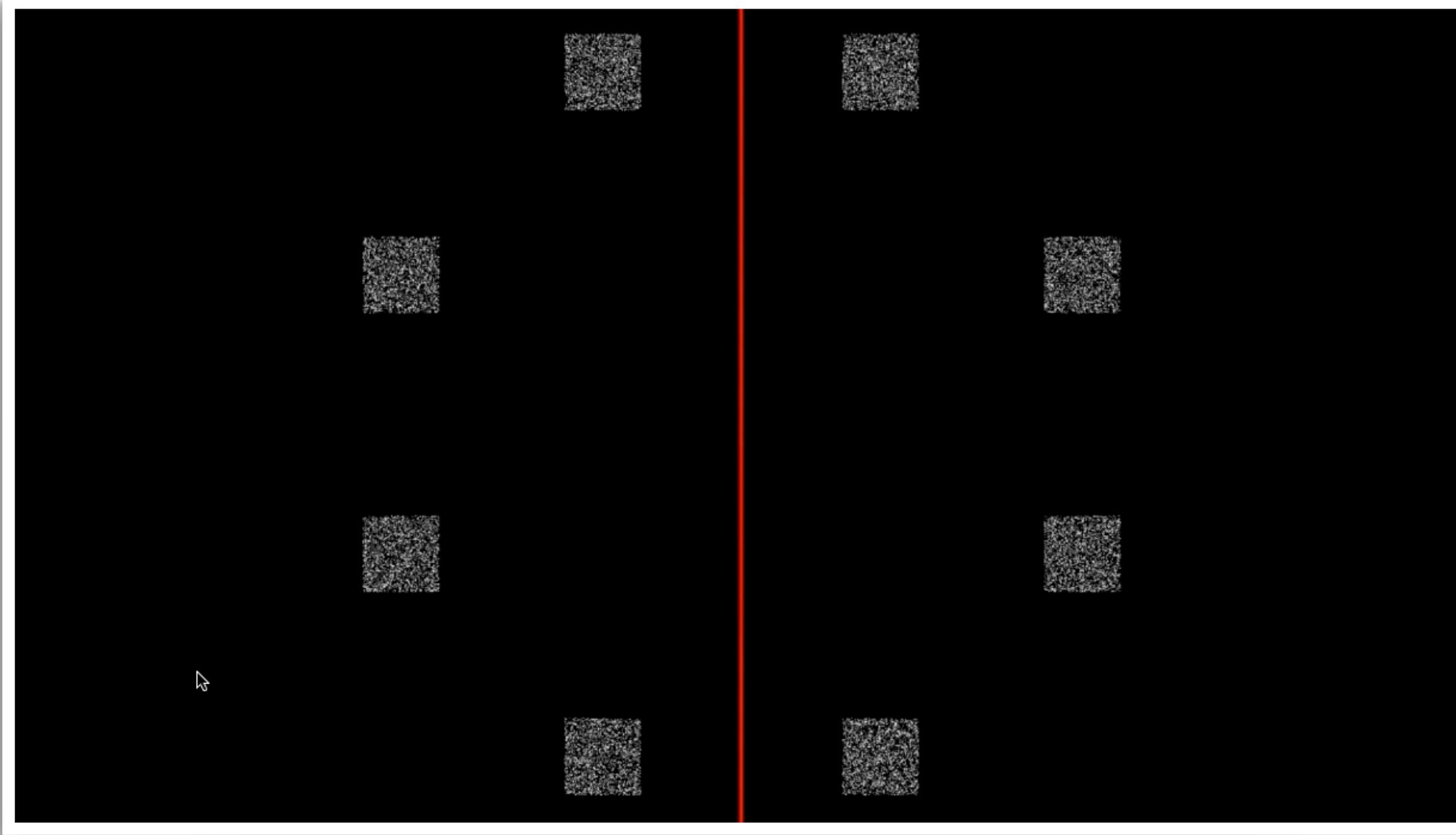
This talk: a detailed analysis of the above question



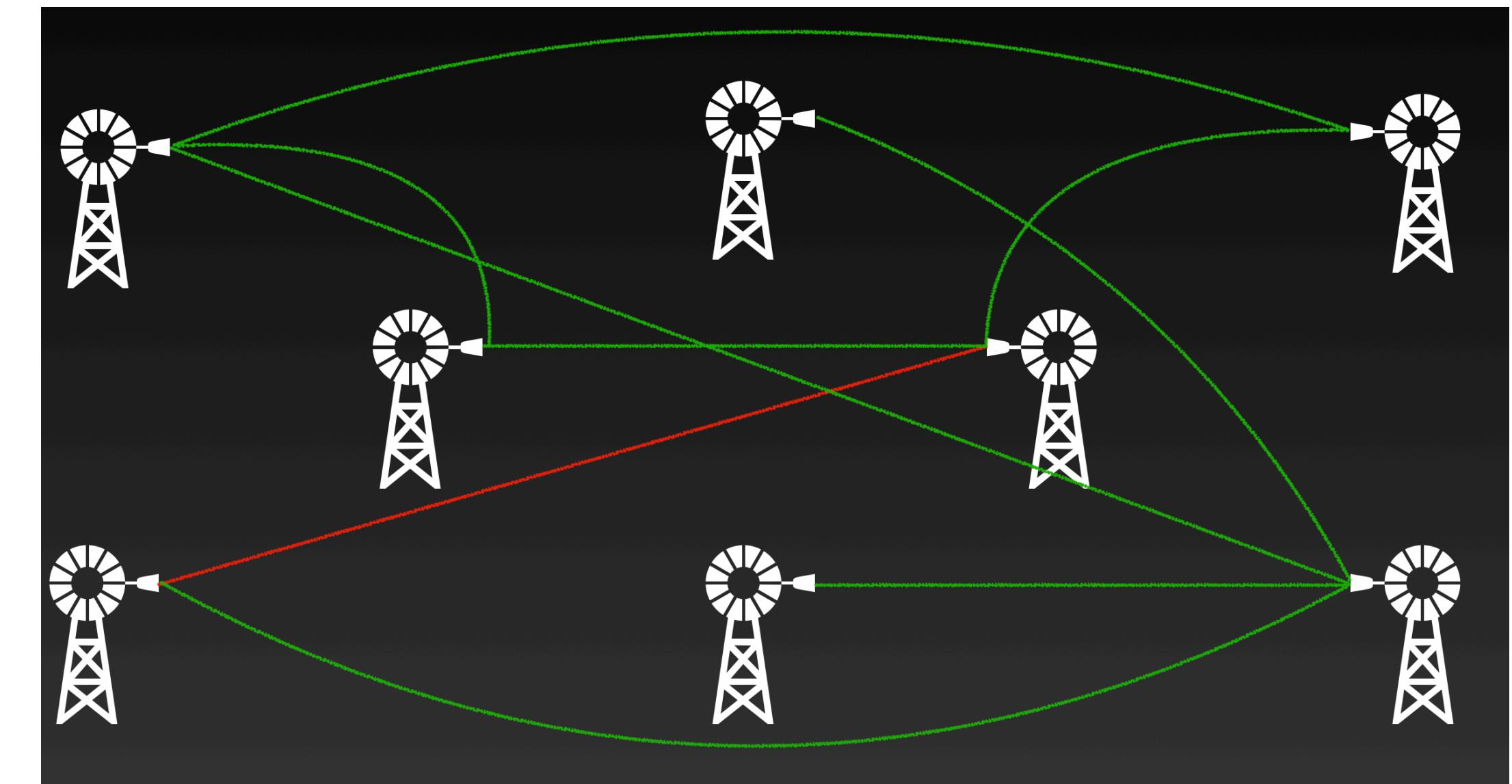
Faulty power line (power systems)

Other Example Applications

“Odd” movie (visual neuroscience)



- A/B testing
- Search for free channel in a cognitive radio network



At the top of search results for "Amazon" was a bad ad, trying to tricking users into falling for a tech support scam.

 By Zack Whittaker for Zero Day | March 16, 2018 -- 00:19 GMT (05:49 IST) | Topic: Security

Google search results for "amazon":

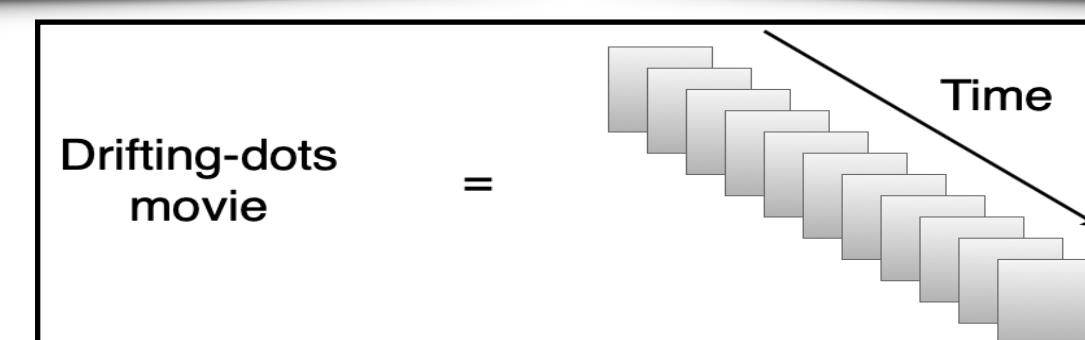
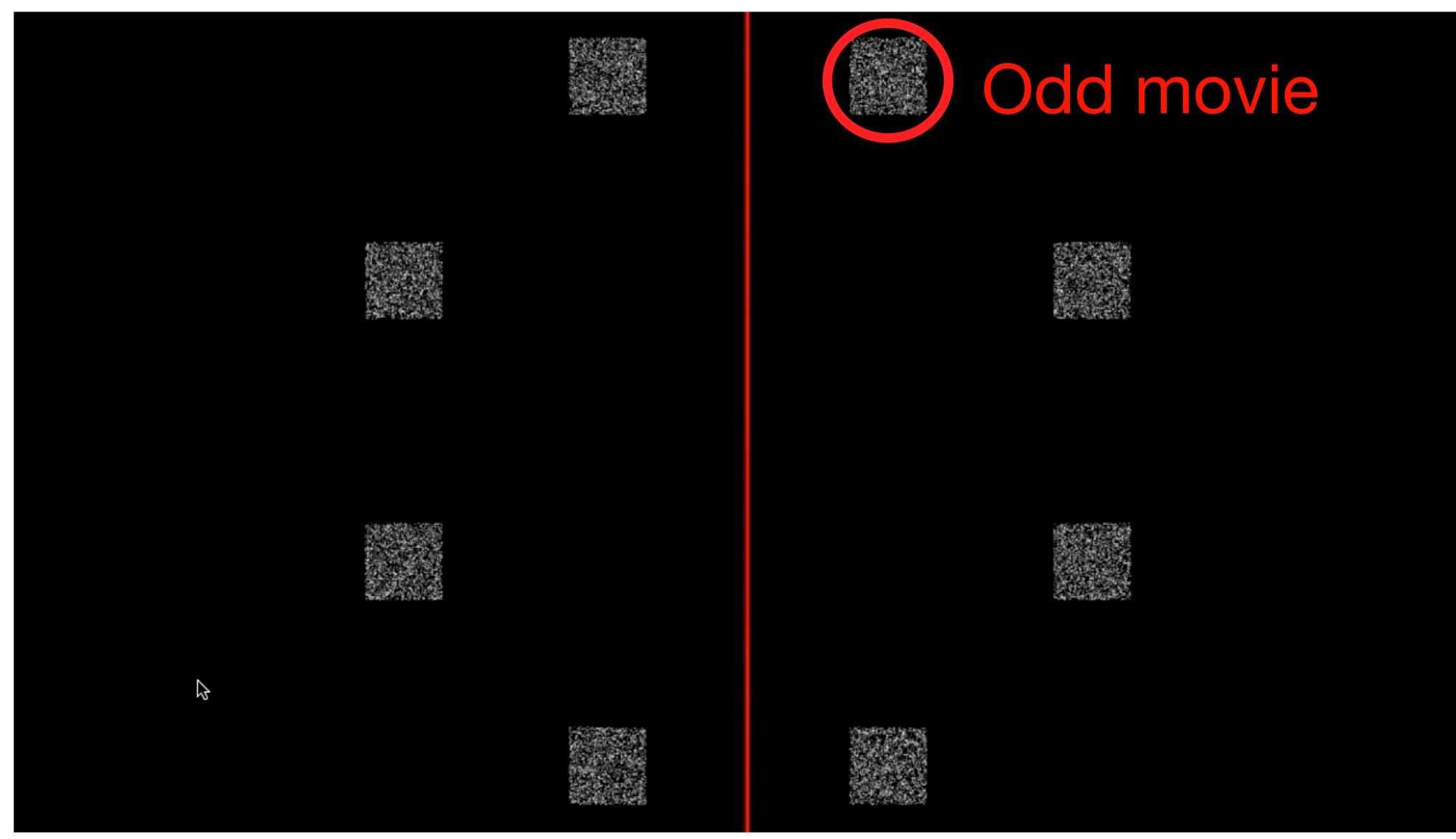
- 1. [Amazon.com | Amazon Official Site | Huge Selection & Great Prices](http://www.amazon.com/)
- 2. [Amazon.com](http://www.amazon.com/)

Sidebar links:

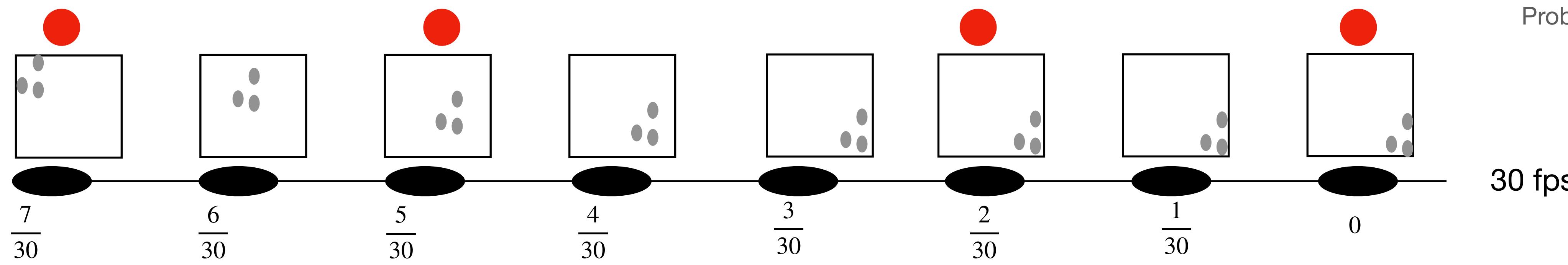
- 5 messages
- f (Facebook)
- in (LinkedIn)
- Twitter
- Email
- Bell (Notifications)
- MORE FROM ZACK WHITTAKER
- Security: Online security 101: Tips for protecting your privacy from hackers and spies
- Security: US government's "do not buy" list shuts out Russia, China

Malicious ad (recommender systems)

Goal: To find the anomalous Markov arm quickly and accurately



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



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

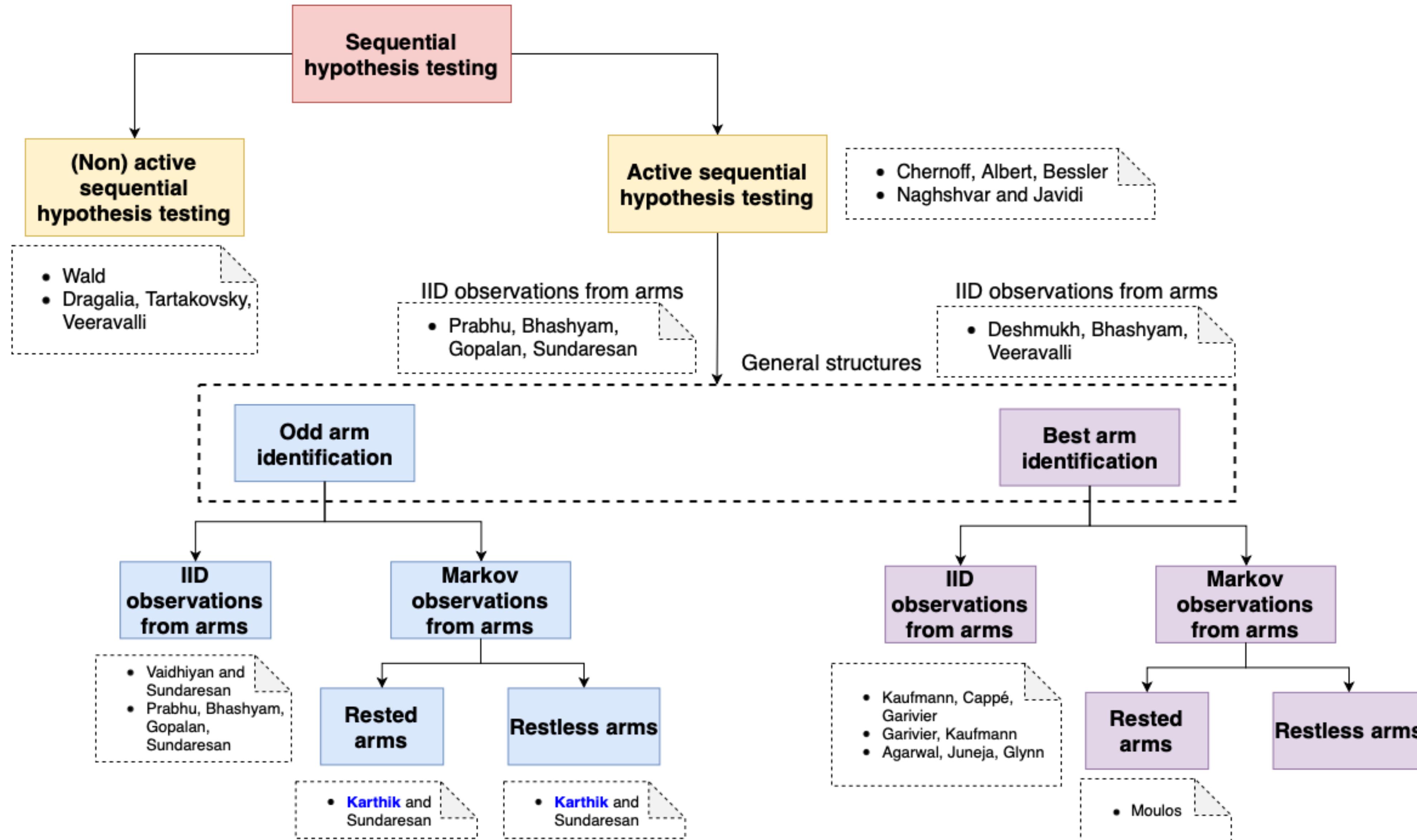
Characterise: $\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) | P_1, P_2]}{\log(1/\epsilon)}$

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**)

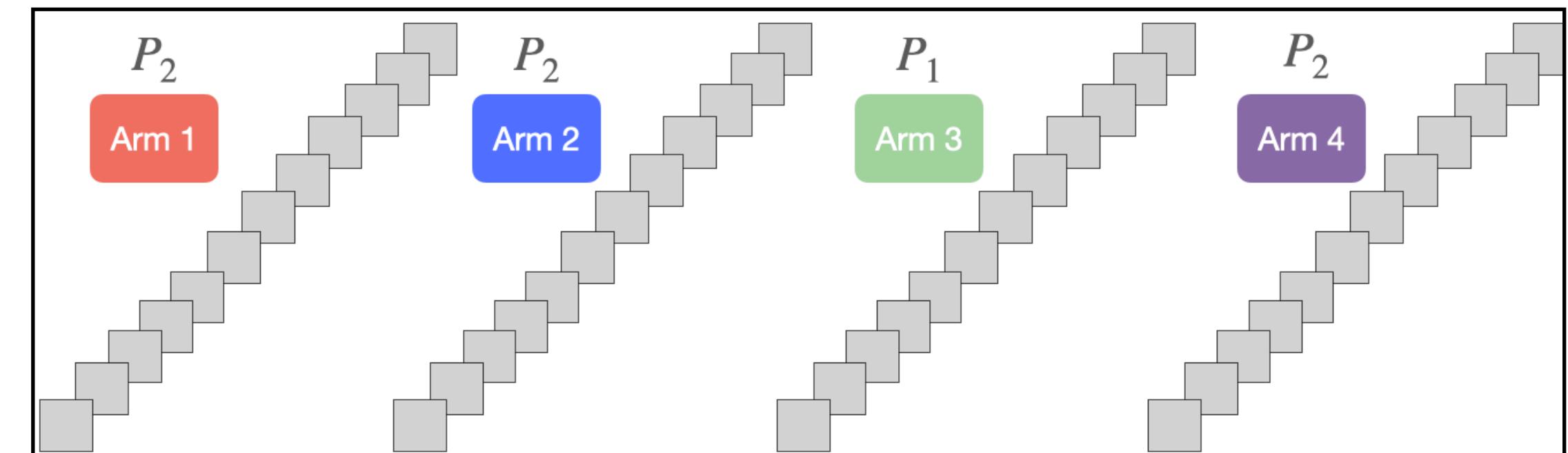
Our Contributions

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

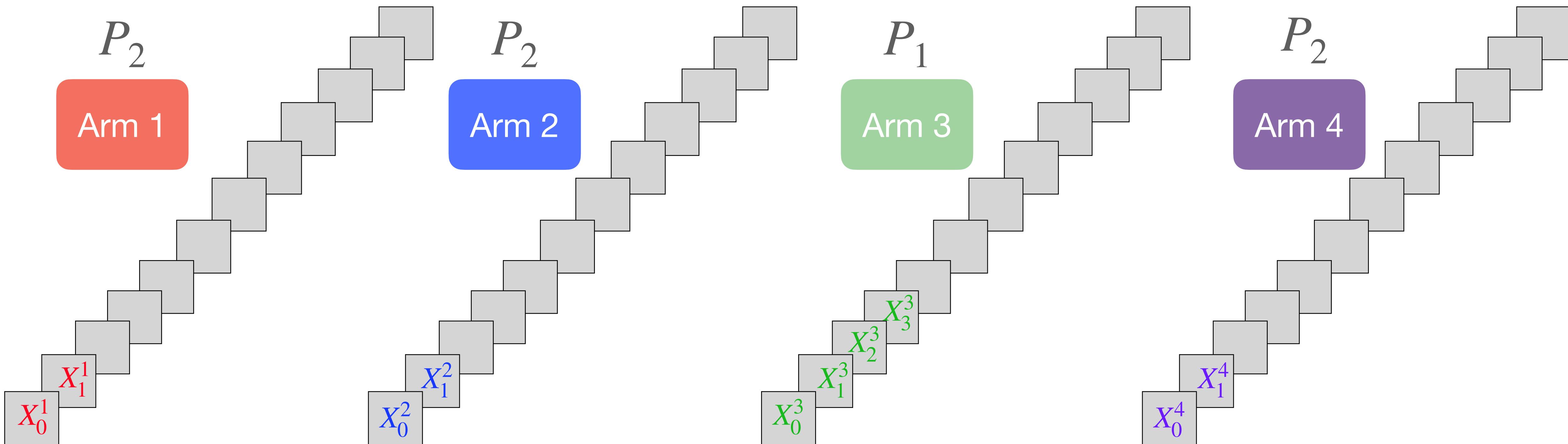
$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) | C]}{\log(1/\epsilon)} \geq \frac{1}{D^*(h, P_1, P_2)}$$

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

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

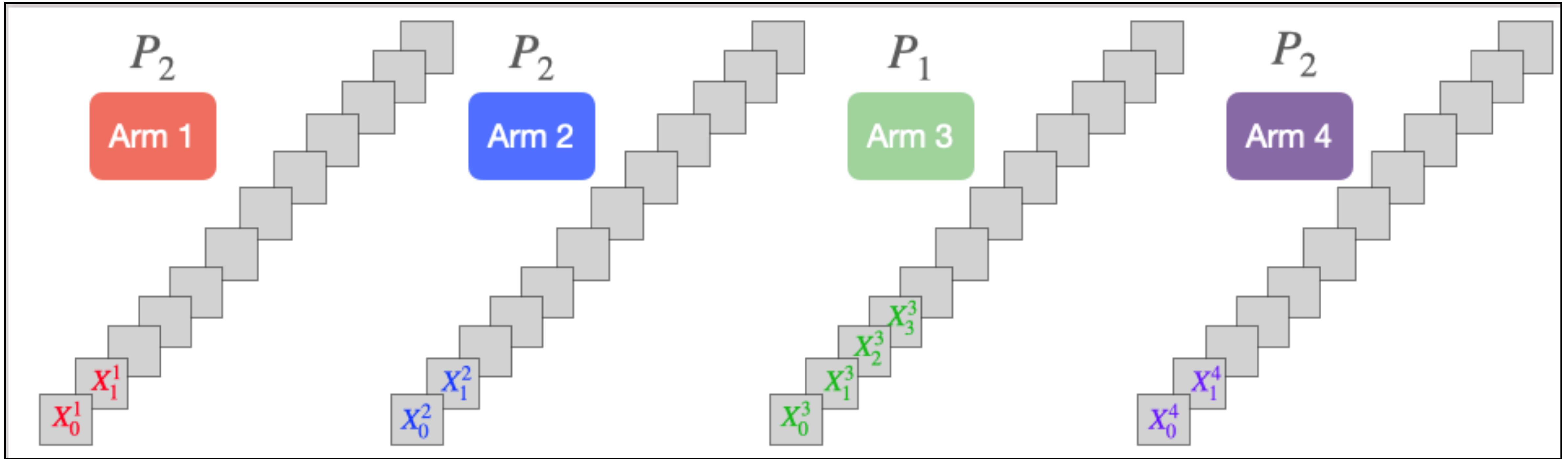


$$D^*(h, P_1, P_2) = \max_{\lambda} \min_{C': h' \neq h} \sum_{a=1}^K \lambda(a) D(P_C^a || P_{C'}^a | \mu_C^a)$$



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 : t th observation from arm a

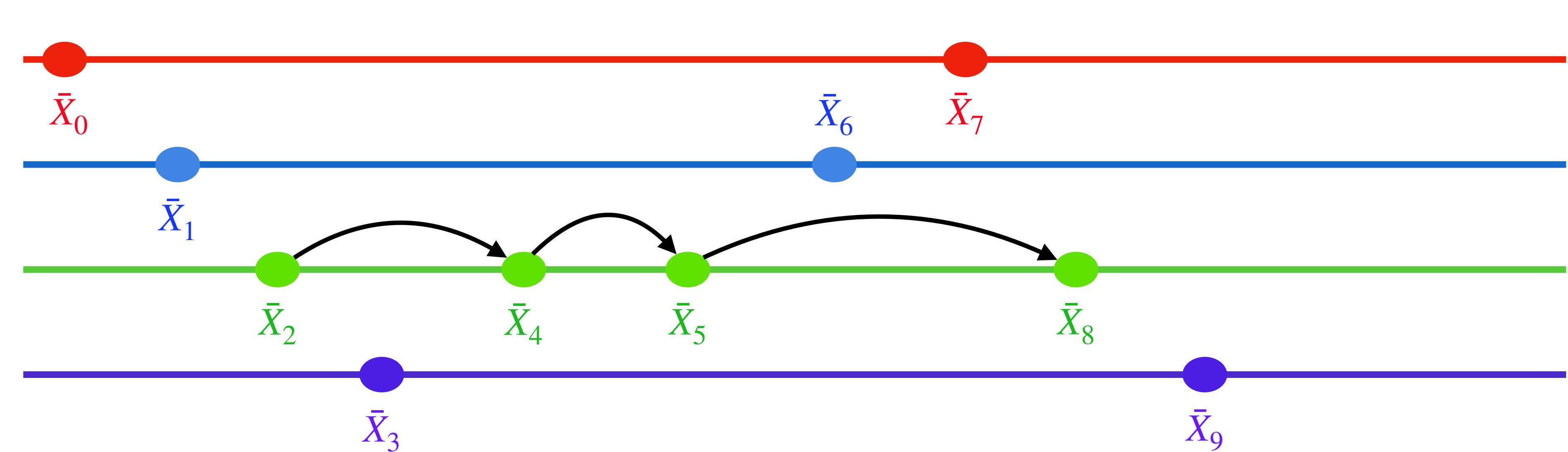
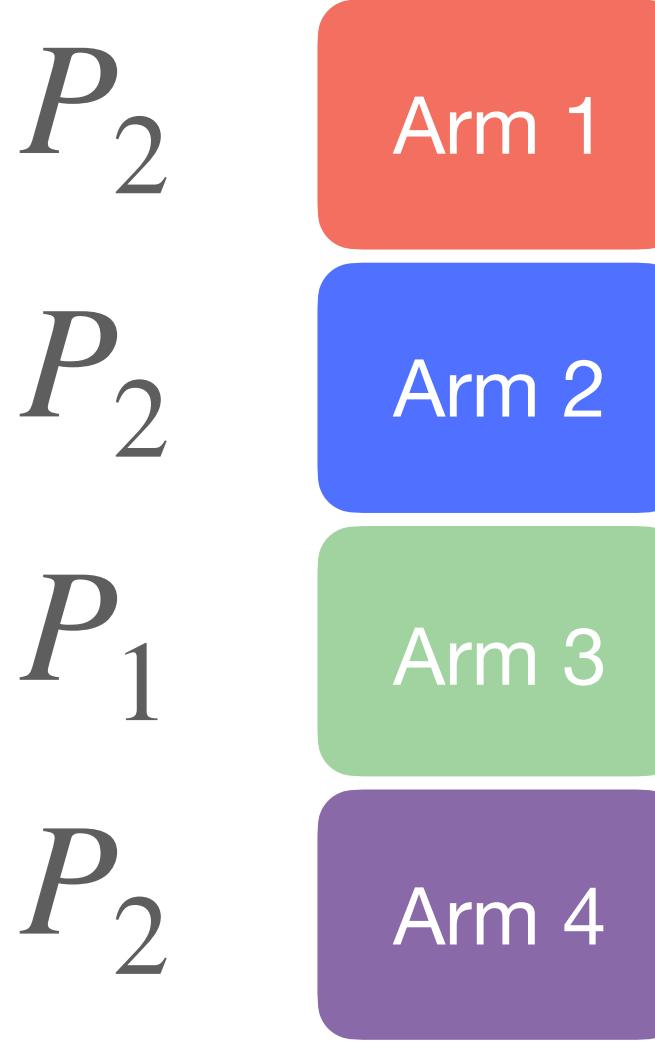


Arm a — sampled $N_a(n)$ times up to time n

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

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

$$Z_C(n) = \log P(A_0, \dots, A_n, X_0^1, \dots, X_{N_1(n)-1}^1, \dots, X_0^K, \dots, X_{N_K(n)-1}^K \mid C)$$



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

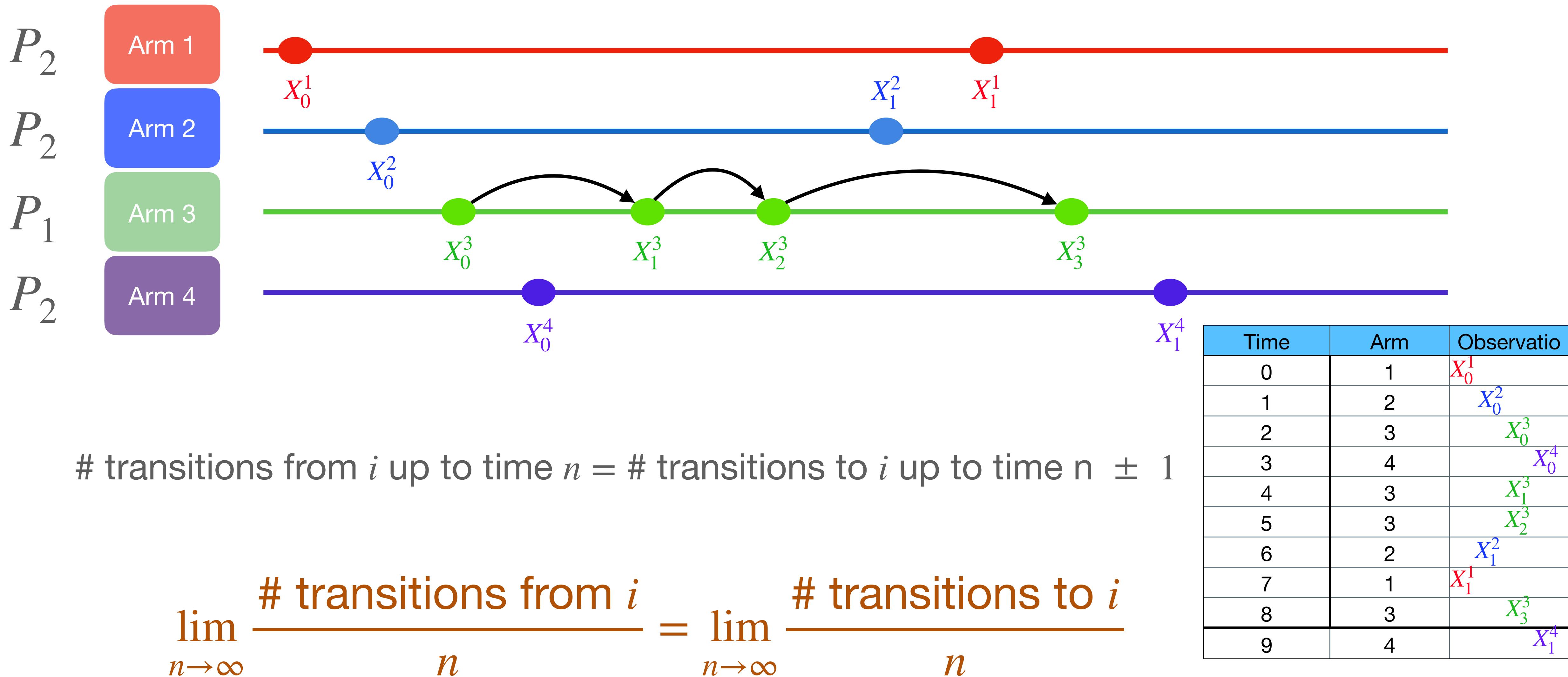
$$Z_C(n) = \sum_{i,j \in \mathcal{S}} N_h(n, i, j) \log P_1(j | i) + \sum_{a \neq h} \sum_{i,j \in \mathcal{S}} N_a(n, i, j) \log P_2(j | i) + \log P_C(A_0) + \log \nu(\bar{X}_0) + \sum_{t=1}^n \log P_C(A_t | A_0, \dots, A_{t-1}, \bar{X}_0, \dots, \bar{X}_{t-1})$$

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

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

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



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

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

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

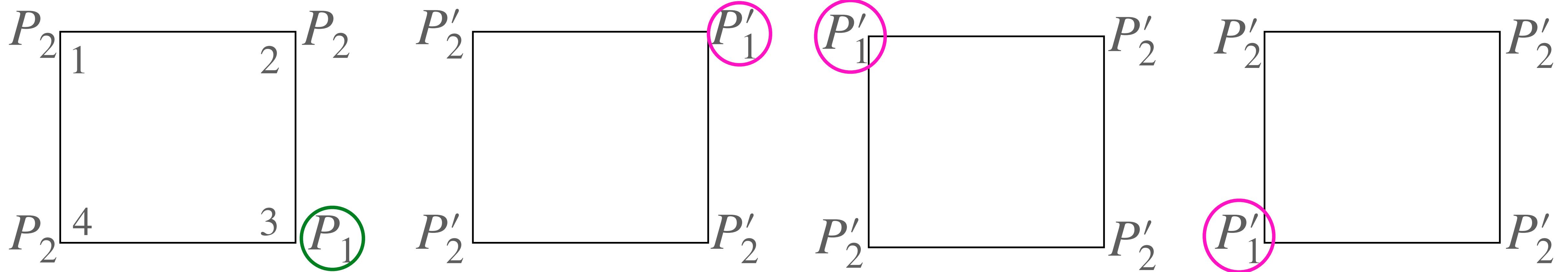
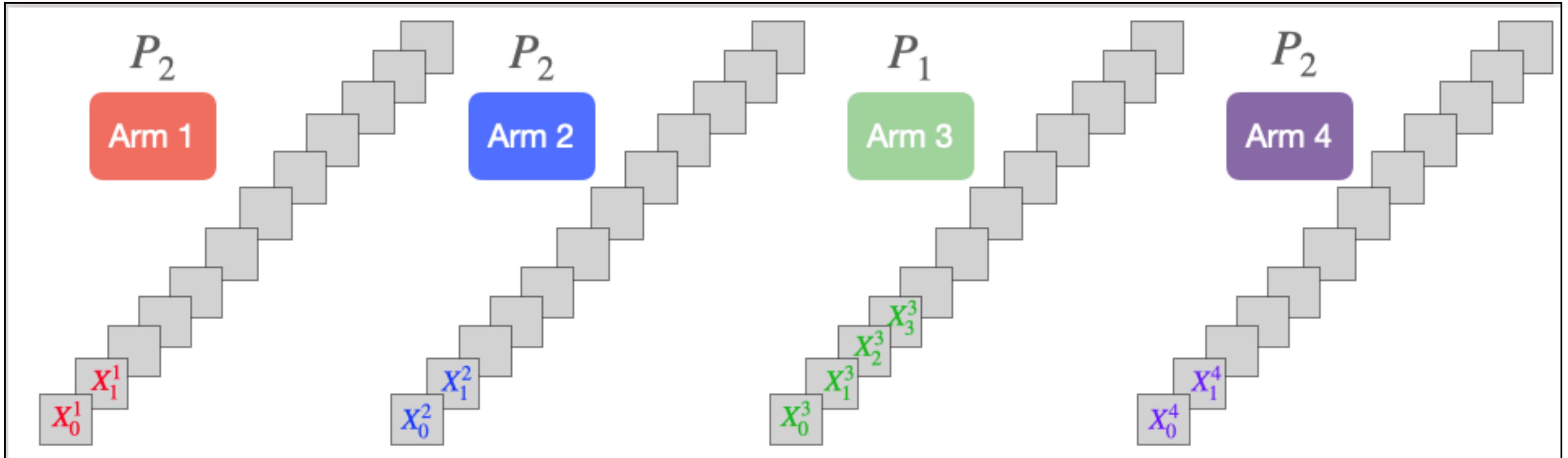
$$\pi \in \Pi(\epsilon)$$

Configuration	Decision = h	Decision = h'	Others
$C = (h, P_1, P_2)$	$\geq 1 - \epsilon$	$\leq \epsilon$	$\leq \epsilon$
$C' = (h', P'_1, P'_2)$	$\leq \epsilon$	$\geq 1 - \epsilon$	$\leq \epsilon$

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

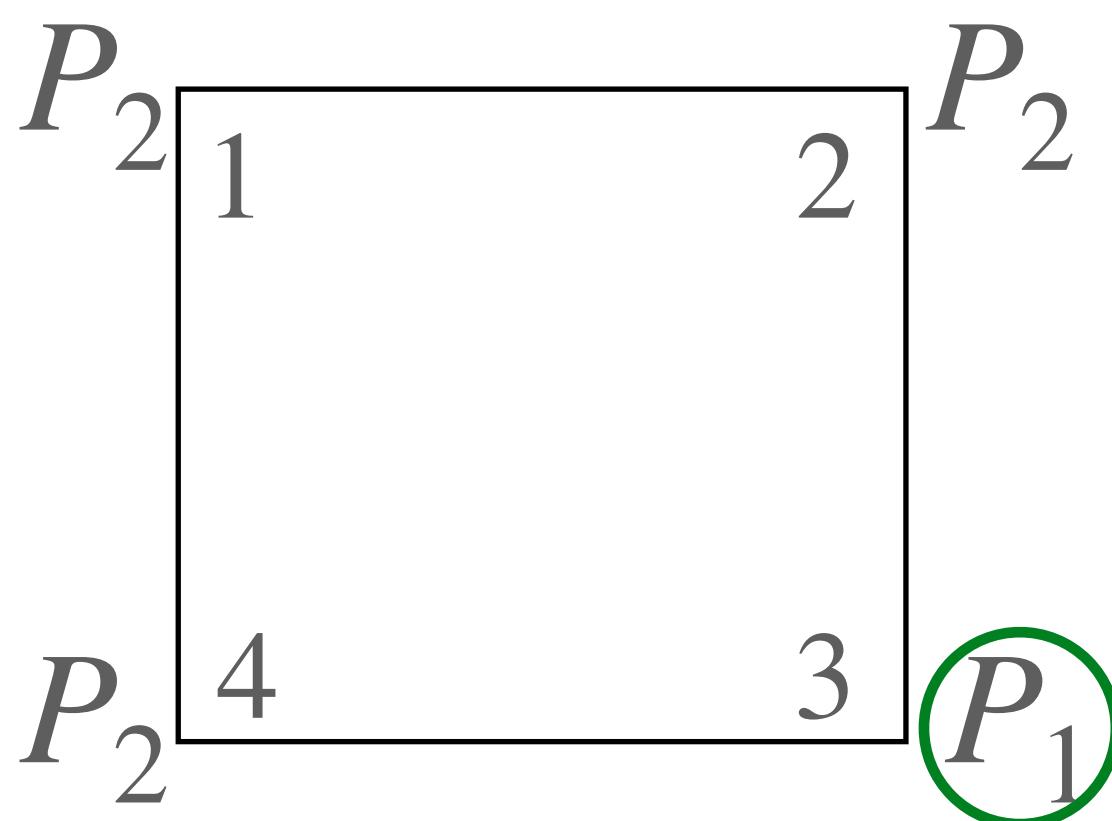
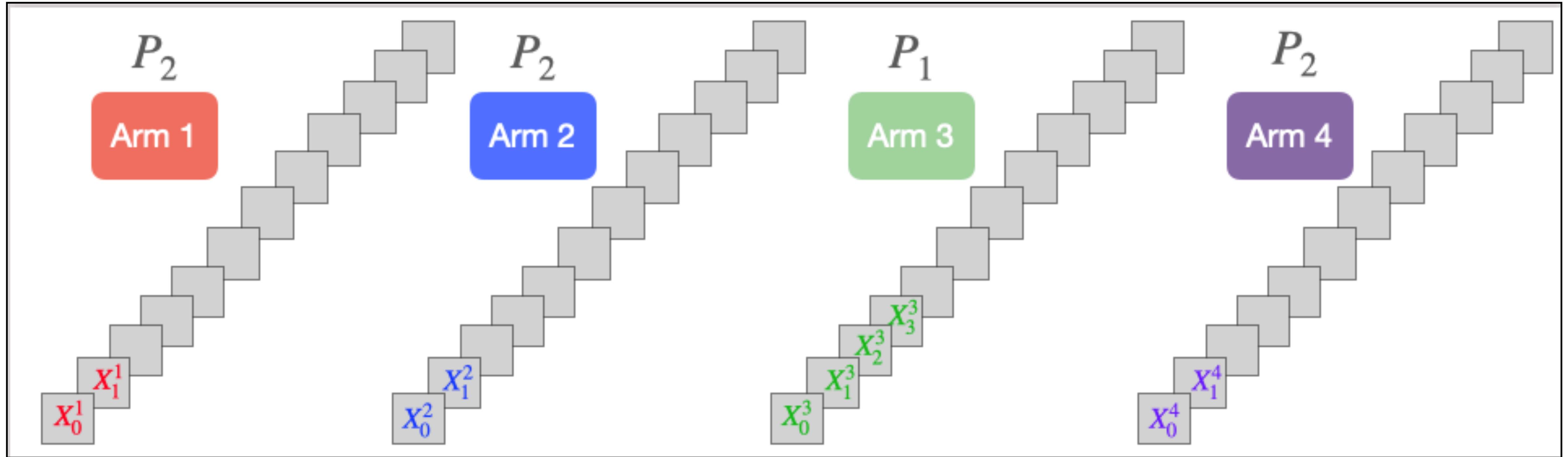
$$C' = (h', P'_1, P'_2) \quad h' \neq h$$

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



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

$$C' = (h', P'_1, P'_2) \quad h' \neq 3$$



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

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)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Select each arm once ($n = 0, \dots, K - 1$)
- 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

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

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

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

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

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

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

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

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

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

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

- Therefore,

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

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

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

Why not Sample the Arms Repeatedly?

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

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

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

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

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

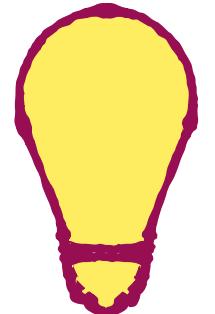
$$\hat{h}(n) \in \arg \max_h \min_{h' \neq h} M_{hh'}(n)$$

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

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

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

Takeaway – Rested Arms with Unknown TPMs



Look for invariant quantities

Rested Arms, Unknown TPMs

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

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

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

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

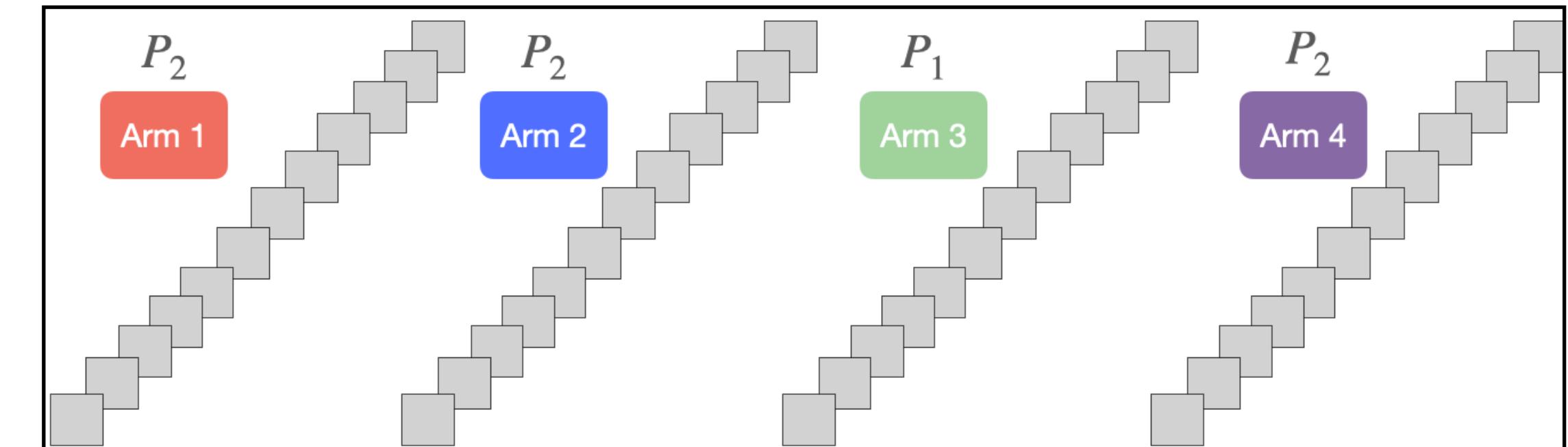
Our Contributions

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

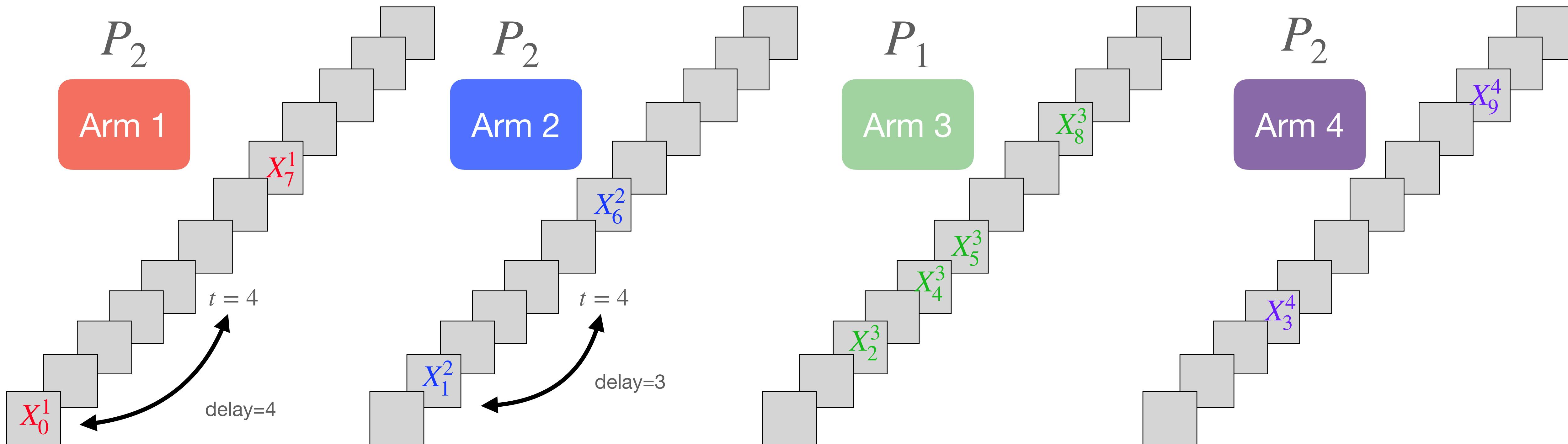
$$\lim_{\epsilon \downarrow 0} \inf_{\pi \in \Pi(\epsilon)} \frac{E[\tau(\pi) | C]}{\log(1/\epsilon)} \geq \frac{1}{R^*(P_1, P_2)}$$

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

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



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



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

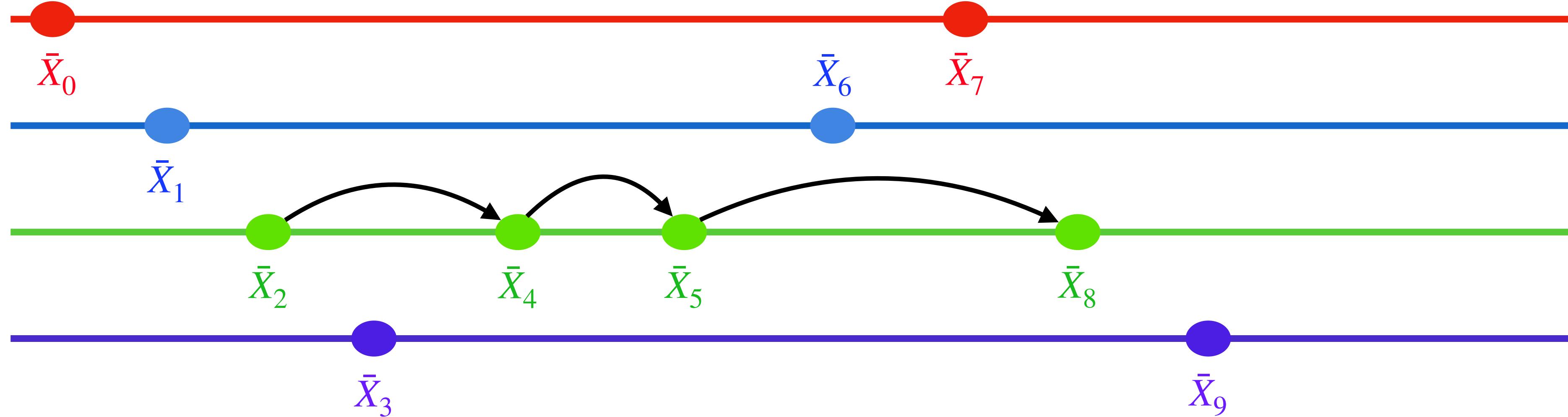
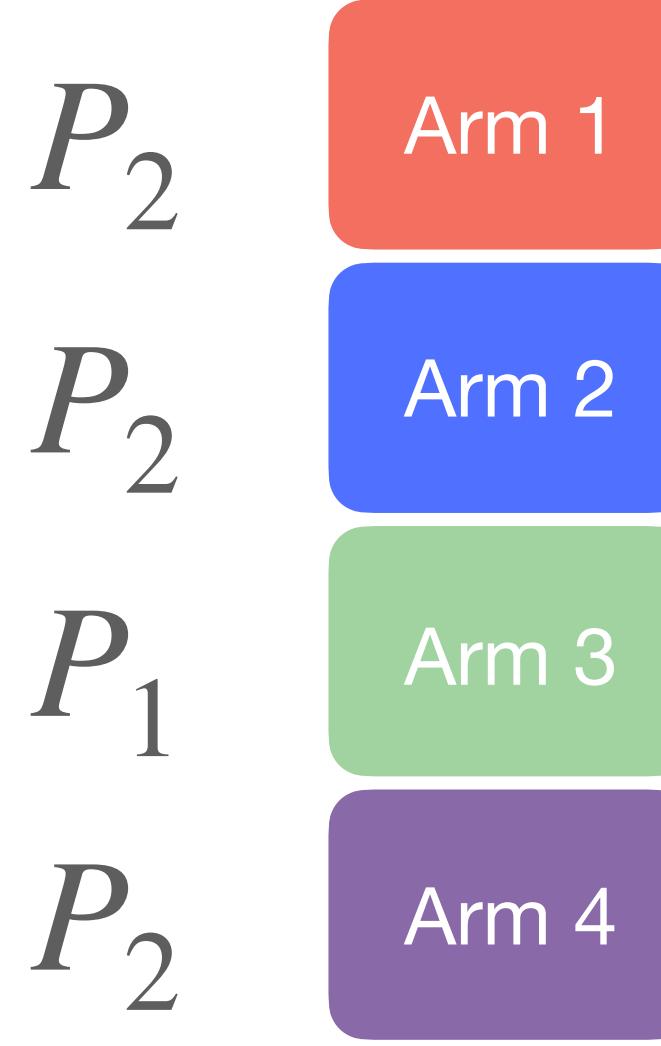
$i_a(t)$: last observed state of arm a at time t

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**:

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

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



If you recall, in the setting of r

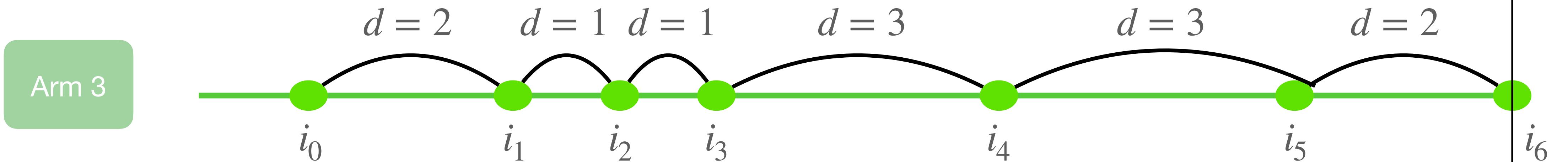
$$C = (h, P_1, P_2)$$

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

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

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

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



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

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

$n = 14$

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

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

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

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

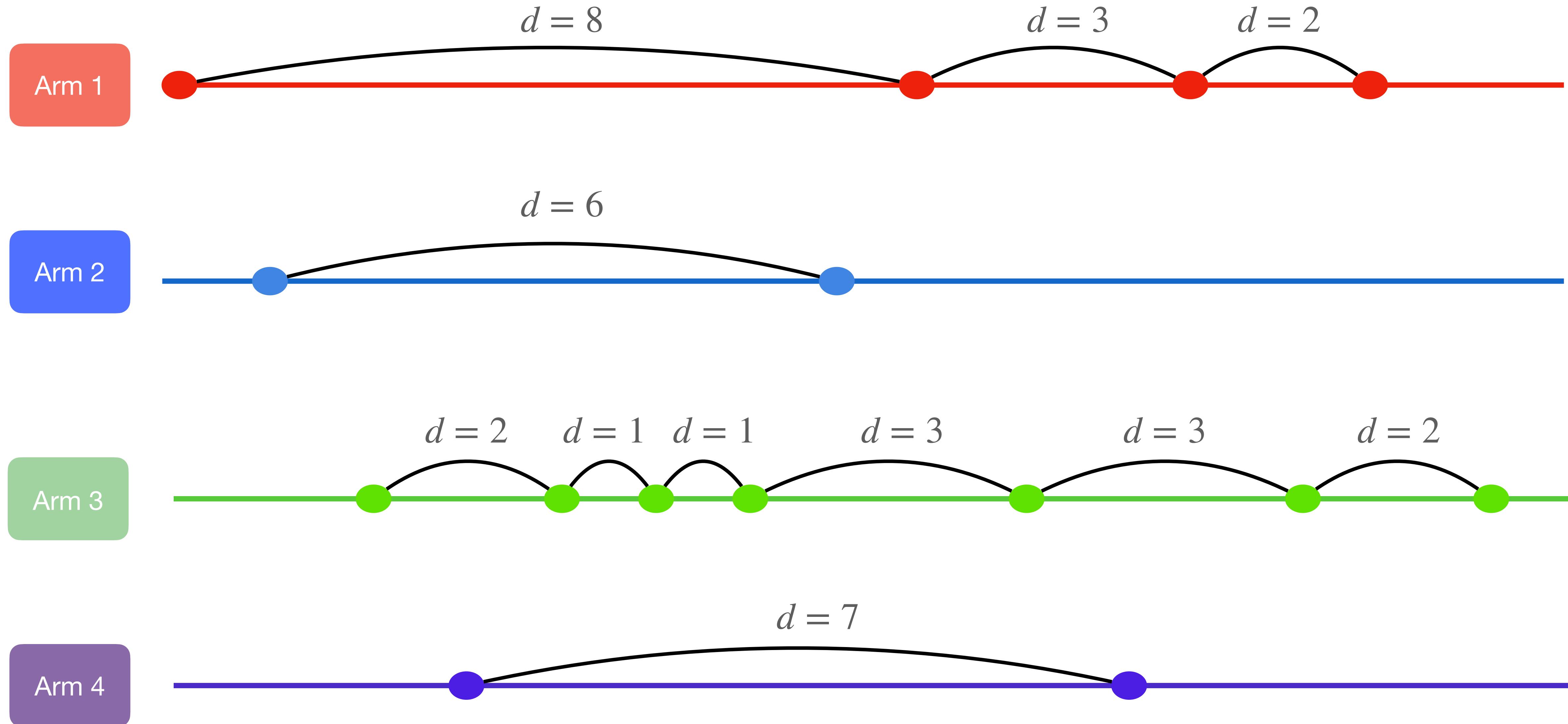
$$\sum_{i \in \mathcal{S}} \sum_{d=1}^{\infty} d \kappa(d, i, a) = 1 \quad \text{for all } a,$$

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

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

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

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



Delays and Last Observed States

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

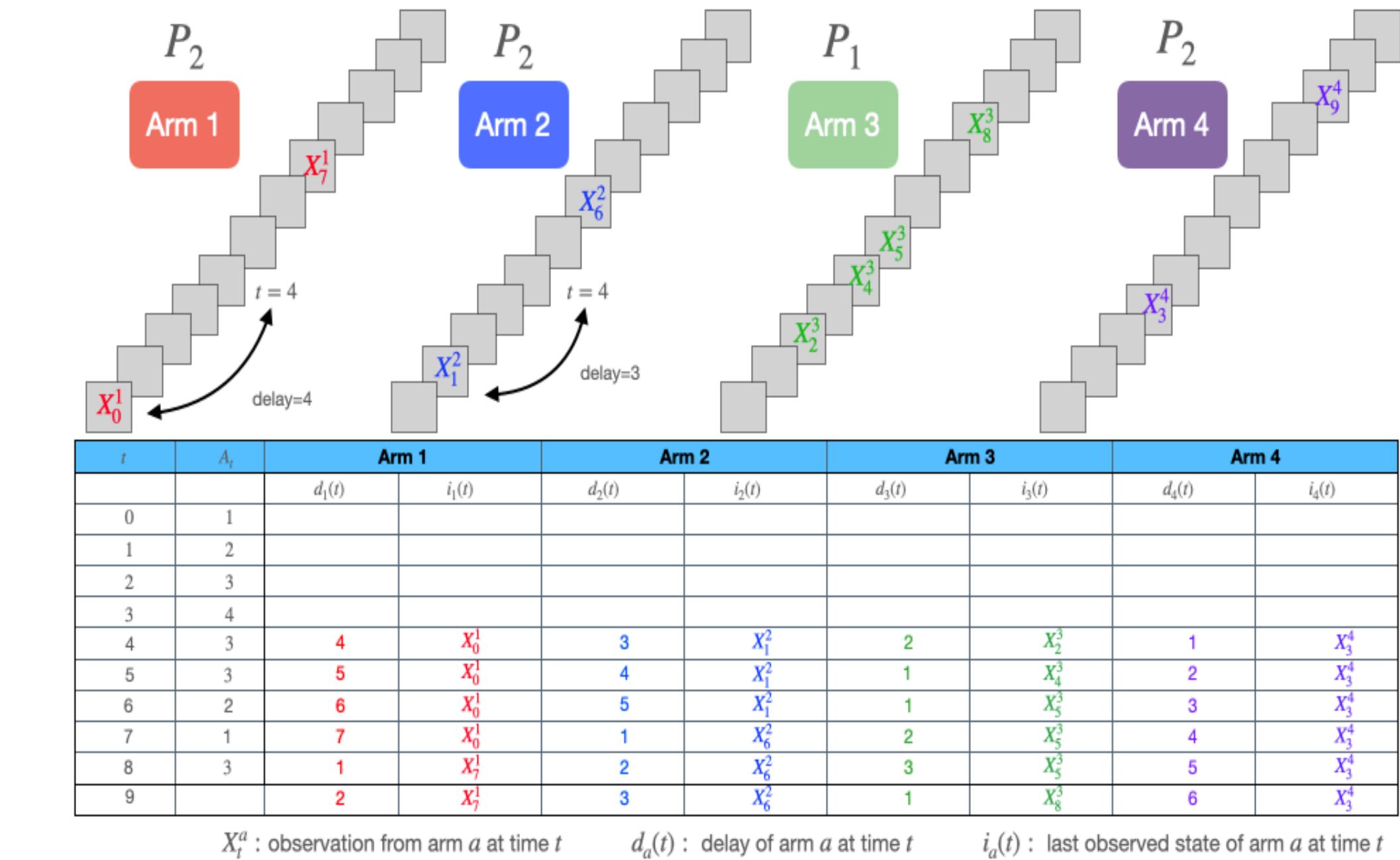
$$P(\underline{d}(t+1), \underline{i}(t+1) | B_0, \dots, B_t, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = P(\underline{d}(t+1), \underline{i}(t+1) | B_t, (\underline{d}(t), \underline{i}(t)))$$

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

Markov Decision Problem (MDP)

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

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



Characterise

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

Markov Decision Problem (MDP)

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

MDP Transition Probabilities

$$\underline{d}(t) = \underline{d} = (4, 3, 2, 1)$$

$$\underline{i}(t) = \underline{i} = (i_1, i_2, i_3, i_4)$$

$$B_t = b$$

$$A_t = 1$$

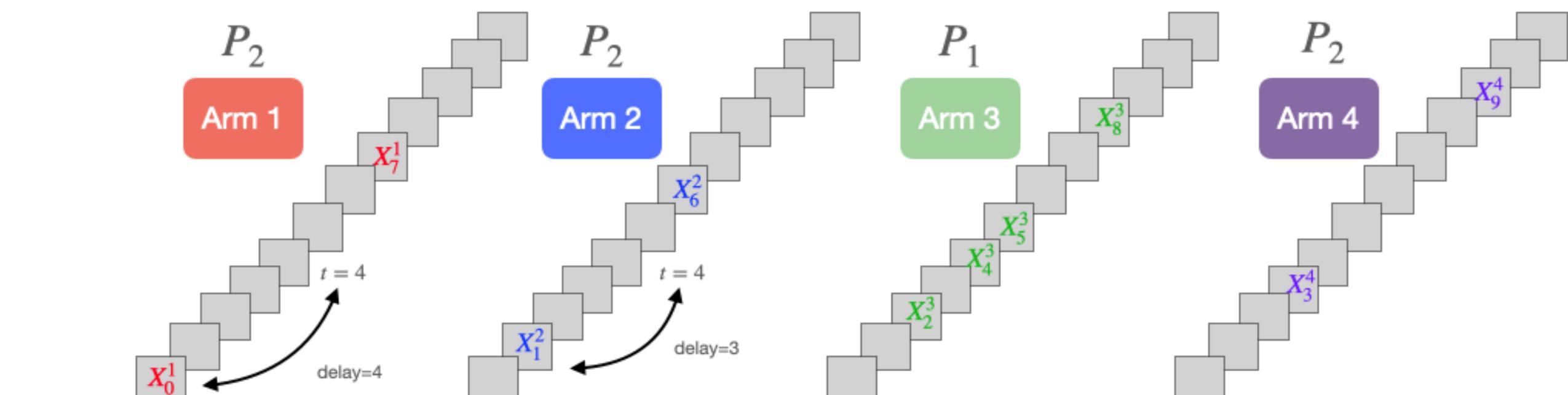
$$P(A_t = 1 | B_t = b) = \frac{\eta}{K} + (1 - \eta) \mathbb{I}_{\{b=1\}}$$

$$\underline{d}(t+1) = \underline{d}' = (1, 4, 3, 2)$$

$$\underline{i}(t+1) = \underline{i}' = (X_t^1, i_2, i_3, i_4)$$

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

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



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

Characterise

$$\liminf_{\epsilon \downarrow 0} \frac{E[\tau(\pi) | C]}{\log(1/\epsilon)}$$

Markov Decision Problem (MDP)

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

MDP Transition Probabilities

$$\underline{d}(t+1) = \underline{d}' = (1, 4, 3, 2)$$

$$\underline{i}(t+1) = \underline{i}' = (X_t^1, i_2, i_3, i_4)$$

$$B_{t+1} = b'$$

$$A_{t+1} = 3$$

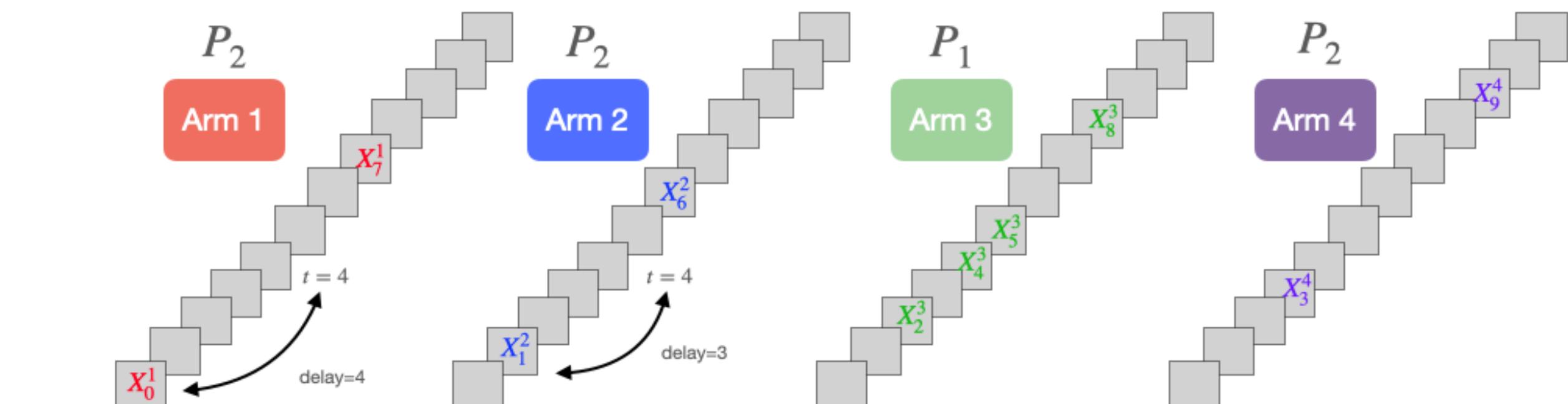
$$P(A_{t+1} = 3 | B_{t+1} = b') = \frac{\eta}{K} + (1 - \eta) \mathbb{I}_{\{b'=3\}}$$

$$\underline{d}(t+2) = \underline{d}'' = (2, 5, 1, 3)$$

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

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

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



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

Characterise

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

Data processing
inequality

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

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

Information theoretic bottleneck:
maximum discrimination per unit time

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

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

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

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

Configuration	Decision = h	Decision = h'	Others
$C = (h, P_1, P_2)$	$\geq 1 - \epsilon$	$\leq \epsilon$	$\leq \epsilon$
$C' = (h', P'_1, P'_2)$	$\leq \epsilon$	$\geq 1 - \epsilon$	$\leq \epsilon$

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

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

SRS Policy

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

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

$$P(B_t | B_0, \dots, B_{t-1}, \{(\underline{d}(s), \underline{i}(s)) : K \leq s \leq t\}) = \lambda(B_t | \underline{d}(t), \underline{i}(t))$$

- 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

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

Difficult to show that
this supremum is
attained



Theorem 8.8.2, Puterman³

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

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

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

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

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 $\pi_1^*(L, \delta)$

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

- Select each arm once ($n = 0, \dots, K - 1$)
- 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 $\lambda_{\hat{h}(n), P_1, P_2, \delta}^*(\cdot | \underline{d}(n), \underline{i}(n))$

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

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

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

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

- Stops in finite time w.p. 1
- $\hat{h}(n) = h$ for all n large, almost surely

- If $L = 1/\epsilon$, then $\pi_1^*(L, \delta) \in \Pi(\epsilon)$

- Upper bound:

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

- Therefore,

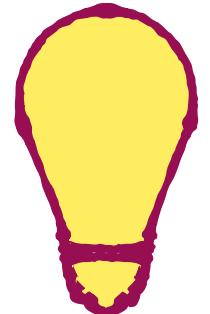
$$\lim_{\delta \downarrow 0} \limsup_{L \rightarrow \infty} \frac{E[\tau(\pi_1^*(L, \delta)) \mid C]}{\log L} \leq \frac{1}{R^*(P_1, P_2)}$$

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

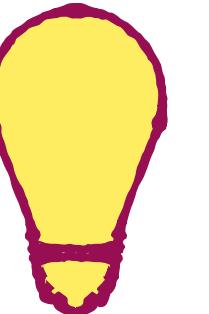
$$\liminf_{\epsilon \downarrow 0} \frac{E[\tau(\pi) \mid C]}{\log(1/\epsilon)} = \frac{1}{R^*(P_1, P_2)}$$

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

Takeaway – Restless Arms with Known TPMs



Look for invariant quantities



Solve a simpler model
and lift the ideas

Rested Arms, Unknown TPMs

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

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

Restless Arms, Known TPMs

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

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

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

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

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

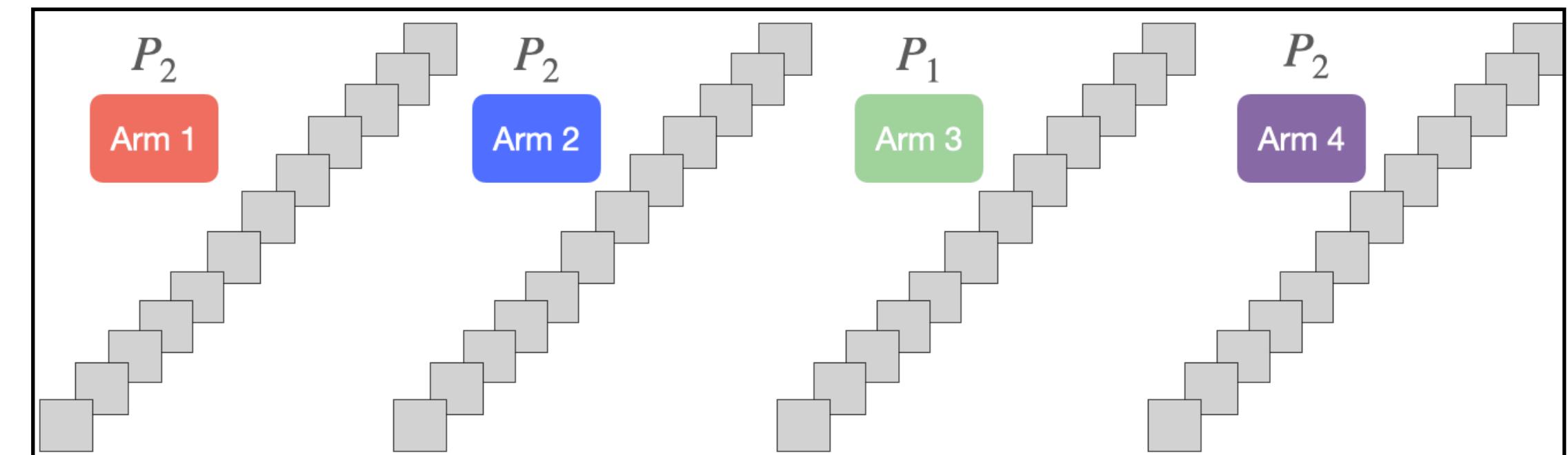
Our Contributions

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

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

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

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



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

Markov Decision Problem (MDP)

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

MDP Transition Probabilities

$$\underline{d}(t) = \underline{d} = (4, 3, 2, 1)$$

$$\underline{i}(t) = \underline{i} = (i_1, i_2, i_3, i_4)$$

$$B_t = b$$

$$A_t = 1$$

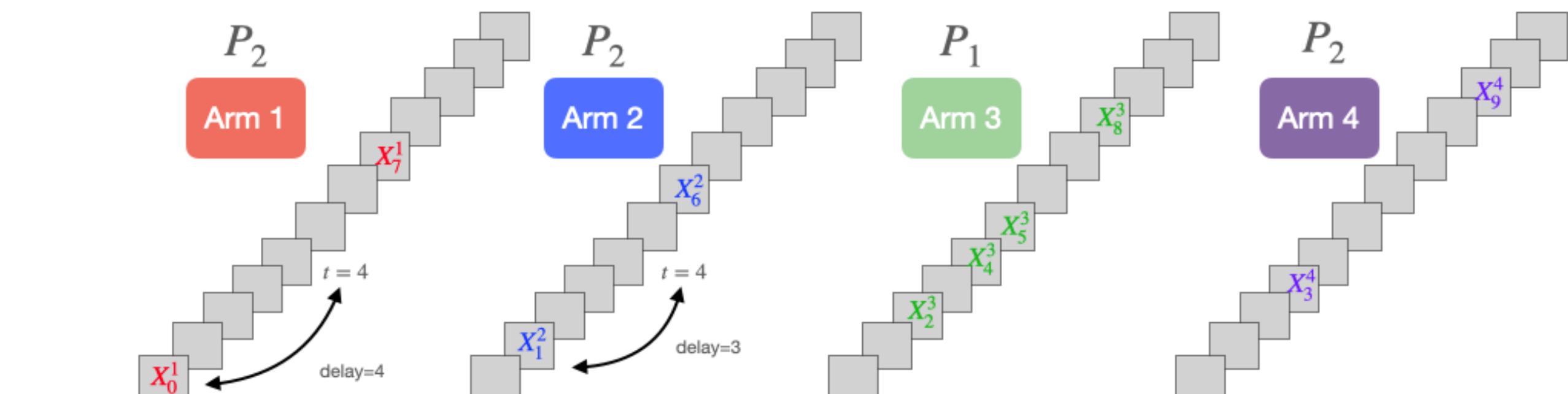
$$P(A_t = 1 | B_t = b) = \frac{\eta}{K} + (1 - \eta) \mathbb{I}_{\{b=1\}}$$

$$\underline{d}(t+1) = \underline{d}' = (1, 4, 3, 2)$$

$$\underline{i}(t+1) = \underline{i}' = (X_t^1, i_2, i_3, i_4)$$

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

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



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

Characterise

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

MDP Transition Probabilities

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

Data processing
inequality

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

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

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

Information theoretic bottleneck:
maximum discrimination per unit time

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

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

$$k_{CC'}(\underline{d}, \underline{i}, a) = \begin{cases} D(P_1^{d_a}(\cdot | i_a) \| (P'_2)^{d_a}(\cdot | i_a)), & a = h, \\ D(P_2^{d_a}(\cdot | i_a) \| (P'_1)^{d_a}(\cdot | i_a)), & a = h', \\ D(P_2^{d_a}(\cdot | i_a) \| (P'_2)^{d_a}(\cdot | i_a)), & a \neq h, h', \end{cases}$$

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

Configuration	Decision = h	Decision = h'	Others
$C = (h, P_1, P_2)$	$\geq 1 - \epsilon$	$\leq \epsilon$	$\leq \epsilon$
$C' = (h', P'_1, P'_2)$	$\leq \epsilon$	$\geq 1 - \epsilon$	$\leq \epsilon$

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

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

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 $\delta > 0$

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

⋮

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

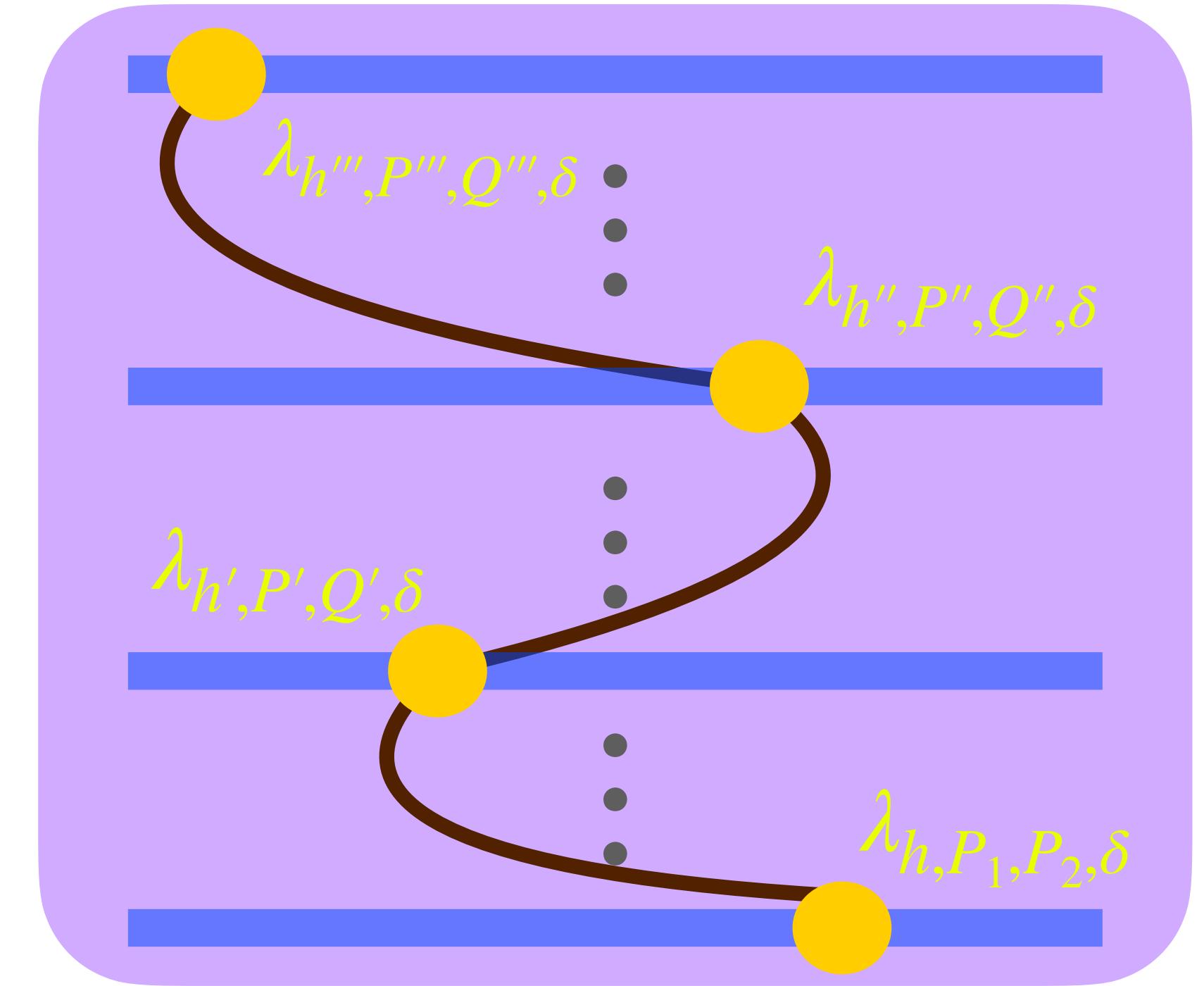
⋮

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

⋮

$$(h, P_1, P_2)$$

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

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

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

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

Arbitrary P, Q

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

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

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

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

- For $n = 0, \dots, K - 1$, 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 $\hat{h}(n) \in \arg \max_h \min_{h' \neq h} \log \frac{\text{avg. likelihood up to time } n \text{ when } h \text{ is the odd arm}}{\text{max. likelihood up to time } n \text{ when } h' \text{ is the odd arm}}$
$$\hat{h}(n) = \arg \max_h \min_{h' \neq h} \log \frac{\sum_{i=1}^n \hat{P}_{h,i}}{\sum_{i=1}^n \hat{P}_{h',i}}$$
 - 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 $\lambda_{\hat{h}(n), \hat{P}_1(n), \hat{P}_2(n), \delta}(\cdot | \underline{d}(n), \underline{i}(n))$
 - Update $n \leftarrow n + 1$

Principle of certainty equivalence

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

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

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

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

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

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

In a Nutshell

Rested Arms, Unknown TPMs

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

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

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

Restless Arms, Known TPMs

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

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

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

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 $\bar{\epsilon}^* \in (0,1)$ such that for any (h, P, Q) , the TPMs $P, Q \in \mathcal{P}(\bar{\epsilon}^*)$

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 $\eta = 0$

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

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

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



What happens as $\eta \downarrow 0$?

The Case $\eta = 0$

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

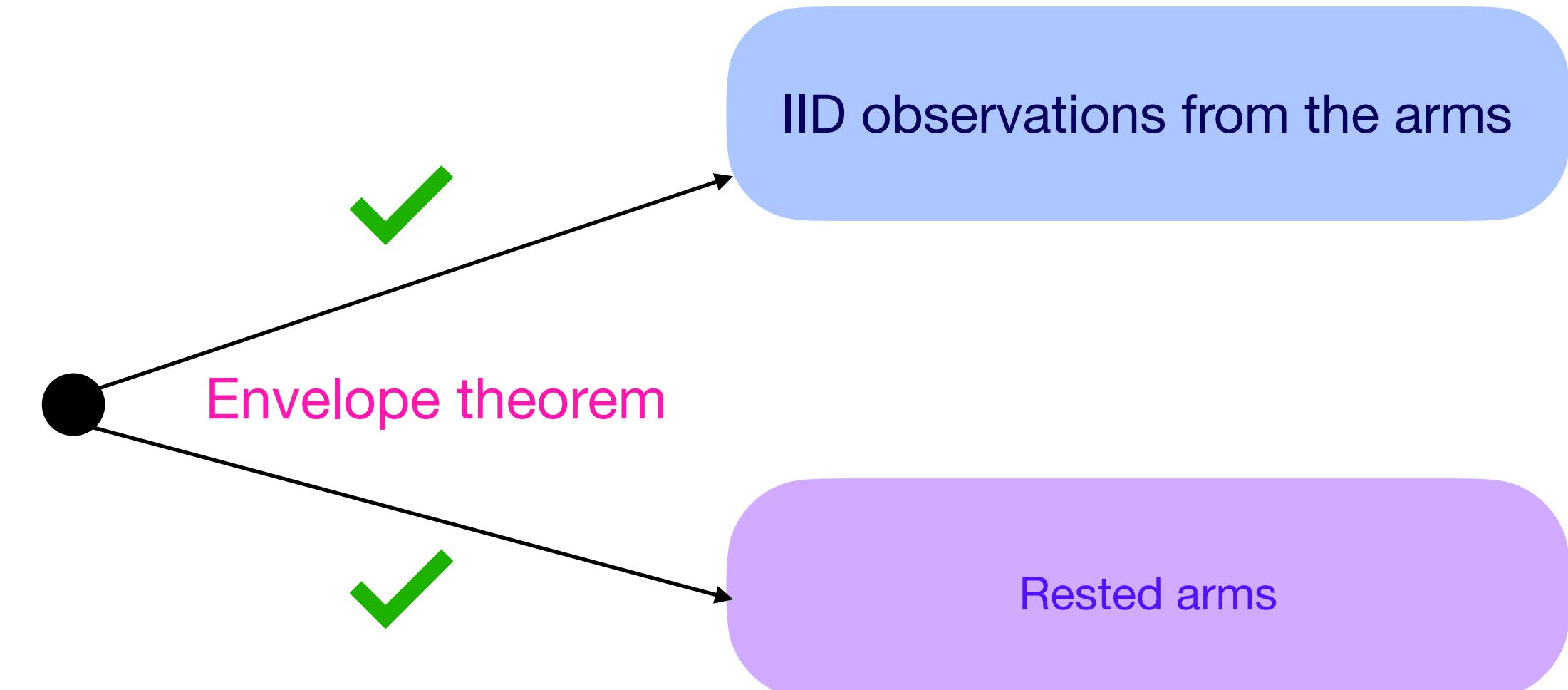
Monotonicity: $\eta' < \eta \implies R_\eta^*(P_1, P_2) \leq R_{\eta'}^*(P_1, P_2)$

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

$$R_0^*(P_1, P_2) = \sup_{\pi^\lambda \in \Pi_{SRS}} \min_{h' \neq h} \sum_{(\underline{d}, \underline{i}) \in \mathbb{S}} \sum_{a=1}^K \mu^\lambda(\underline{d}, \underline{i}) \cdot \lambda(a | \underline{d}, \underline{i}) \cdot k_{hh'}(\underline{d}, \underline{i}, a)$$

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

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



The Case $\eta = 0$

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

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

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

What happens as
 $\eta \downarrow 0$?

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

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

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

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
 $d \in \{1, 2, \dots\}$
- Q-learning for restless bandits⁴
- 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

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

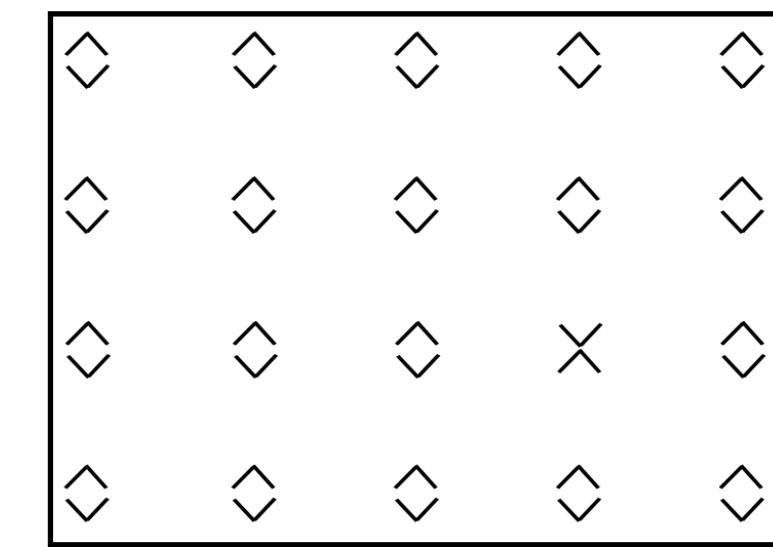
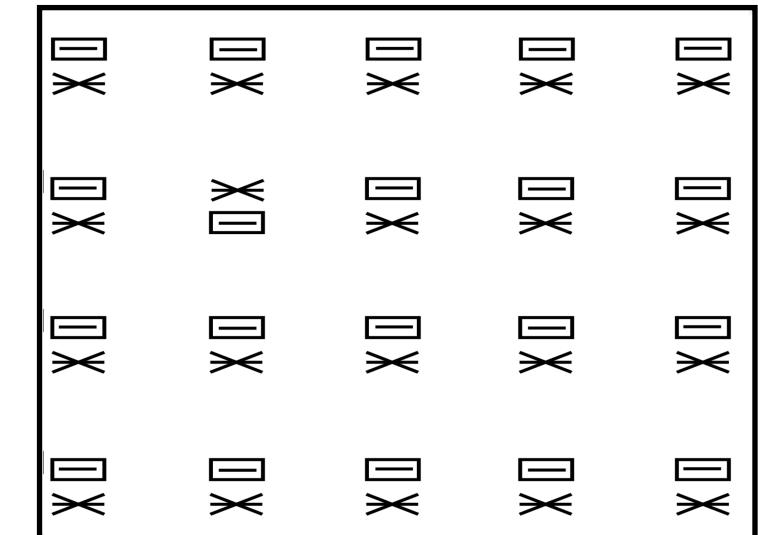
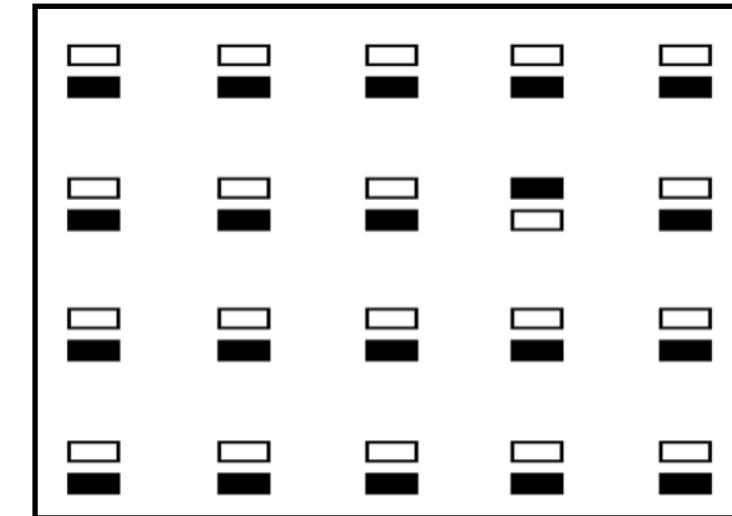
$$\inf_{\pi \in \Pi(\epsilon)} E^\pi[\tau(\pi) | P_1, P_2] \approx \alpha(P_1, P_2) \cdot \log(1/\epsilon)$$

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

Future Work (contd.)

- Switching costs
- Sophisticated visual search models
- General sequential hypothesis testing (L -anomalous arms identification, best arm identification)
- Hidden Markov models

Grasping multiple objects at once
Memory Constrained Search



• Prabhu, Bhashyam,
Gopalan, Sundaresan

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](https://arxiv.org/abs/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



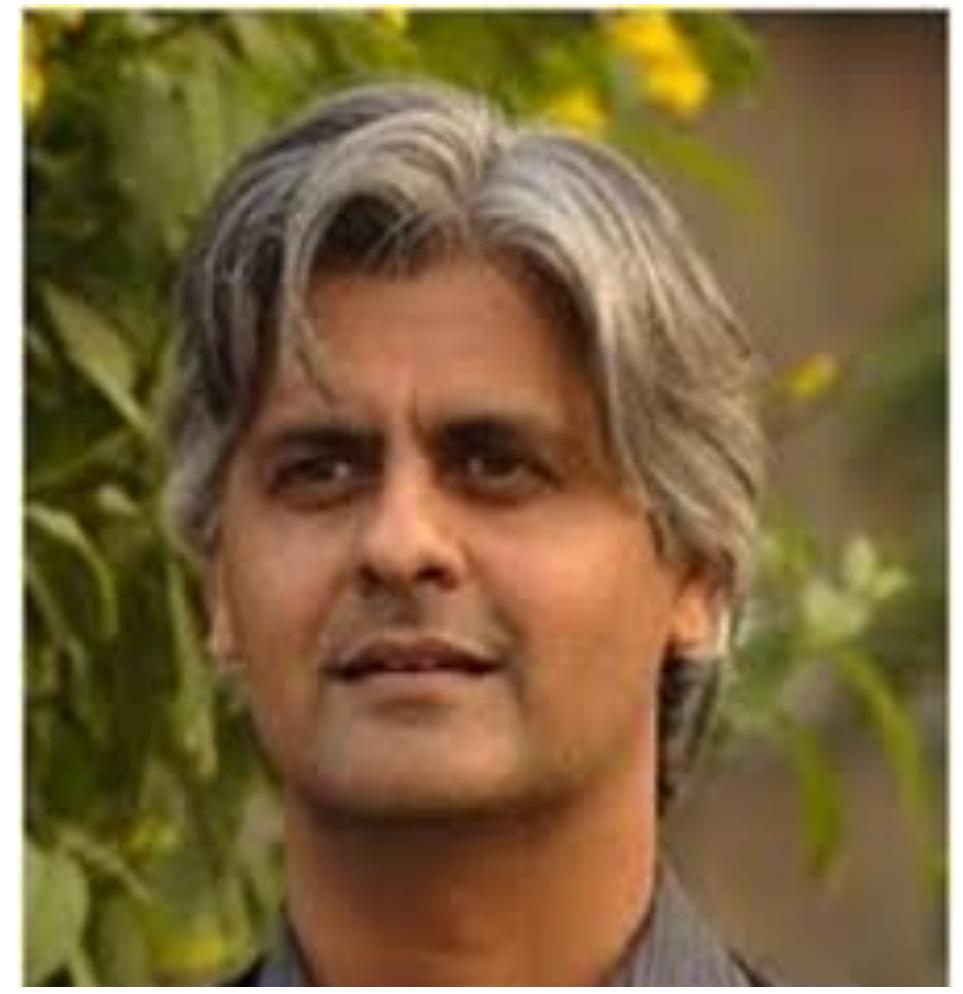
Rajesh Sundaresan
IISc



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



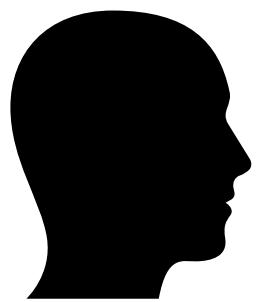
Prathamesh Mayekar
PhD student, IISc



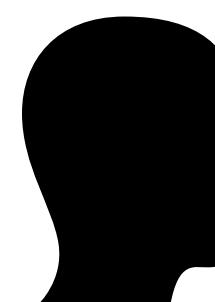
Sahasranand K R
PhD student, IISc



Shubhada Agrawal
PhD student, TIFR Mumbai

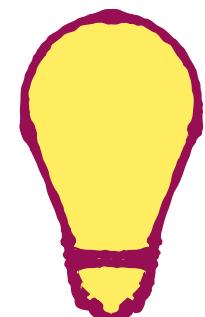


Gayathri Prabhu
PhD student, IIT Madras

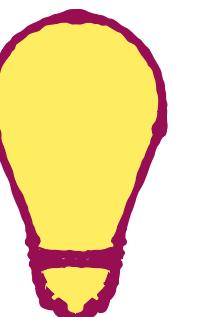


Lakshmi Priya M E
PhD student, IISc

Takeaway

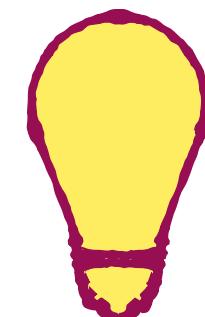


Look for invariant quantities



Solve a simpler model
and lift the ideas

Identifiability



Search for the right keyword

Rested Arms, Unknown TPMs

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

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

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

Restless Arms, Known TPMs

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

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

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

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 $\bar{\epsilon}^* \in (0,1)$ such
that for any (h, P, Q) , the
TPMs $P, Q \in \mathcal{P}(\bar{\epsilon}^*)$

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!

Reach out to me at: pnkarthik1992@gmail.com
<https://www.karthikpn.com>