On Detecting an Anomalous Arm in a Multi-armed Bandit with Markov Observations

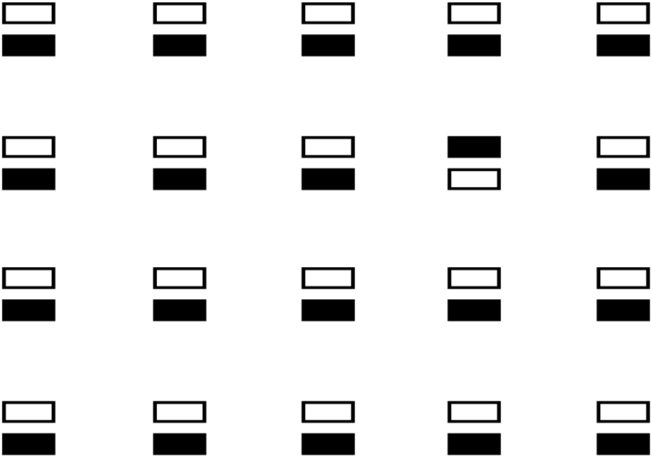
Joint work with Rajesh Sundaresan

P. N. Karthik
Dept. of Electrical Communication Engg.
IISc, Bangalore
periyapatna@iisc.ac.in

STCS Annual Symposium 2020
TIFR, Mumbai
March 1, 2020

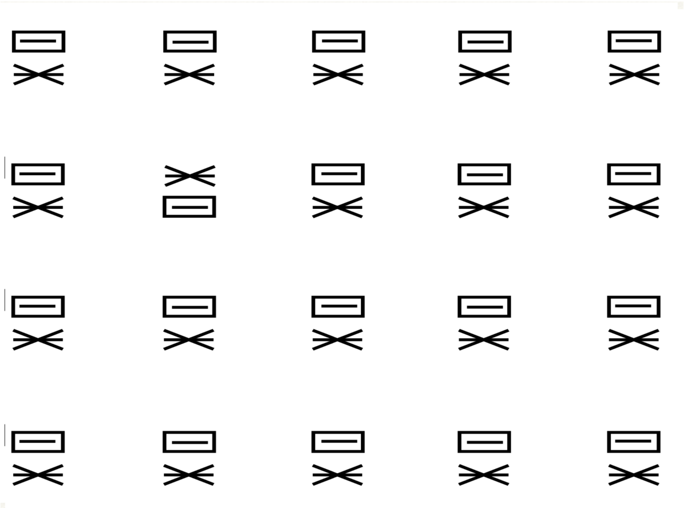
Motivation

A Visual Search Experiment



Can you identify the location of the “odd” image as quickly as possible?

A Visual Search Experiment



Can you identify the location of the “odd” image as quickly as possible?

A Closer Look at the Visual Search Experiment

- In the first example, you were told what the odd and the non-odd images are, but in the second example, you were not

A Closer Look at the Visual Search Experiment

- In the first example, you were told what the odd and the non-odd images are, but in the second example, you were not
- The above information affected the time taken to identify the odd image - shorter time in the first example

A Closer Look at the Visual Search Experiment

- In the first example, you were told what the odd and the non-odd images are, but in the second example, you were not
- The above information affected the time taken to identify the odd image - shorter time in the first example
- Time to identify the odd image $\propto \frac{1}{\text{``closeness'' of the images}}$

A Closer Look at the Visual Search Experiment

- In the first example, you were told what the odd and the non-odd images are, but in the second example, you were not
- The above information affected the time taken to identify the odd image - shorter time in the first example
- Time to identify the odd image $\propto \frac{1}{\text{“closeness” of the images}}$

Can we make precise this notion of closeness?

A Closer Look at the Visual Search Experiment

- In the first example, you were told what the odd and the non-odd images are, but in the second example, you were not
- The above information affected the time taken to identify the odd image - shorter time in the first example
- Time to identify the odd image $\propto \frac{1}{\text{“closeness” of the images}}$

Can we make precise this notion of closeness?

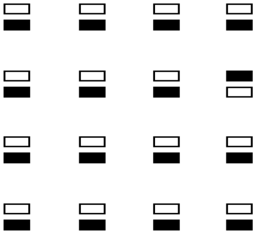
- If the static images in the previous experiments were replaced with dynamic images (movies)[†], what is the notion of closeness of the odd and the non-odd movies?

†Krueger, Paul M., et al. "Evidence accumulation detected in BOLD signal using slow perceptual decision making." Journal of neuroscience methods 281 (2017): 21-32.

Problem Setup: Odd Arm Identification

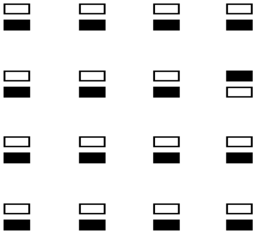
Problem Formulation

- A multi-armed bandit with $K \geq 3$ arms
- Each arm is either an iid process or a Markov process on a finite state space
- The law of one of the arms (the **odd** arm) is P_1, which is different from P_2, the common law of each of the other arms
- P_1 and P_2 may or may not be known
- Sequential arm selections



Problem Formulation

- A multi-armed bandit with $K \geq 3$ arms
- Each arm is either an iid process or a Markov process on a finite state space
- The law of one of the arms (the **odd** arm) is P_1, which is different from P_2, the common law of each of the other arms
- P_1 and P_2 may or may not be known
- Sequential arm selections



Two quantities of interest:

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

(b) Probability of error

Problem Formulation

This talk:

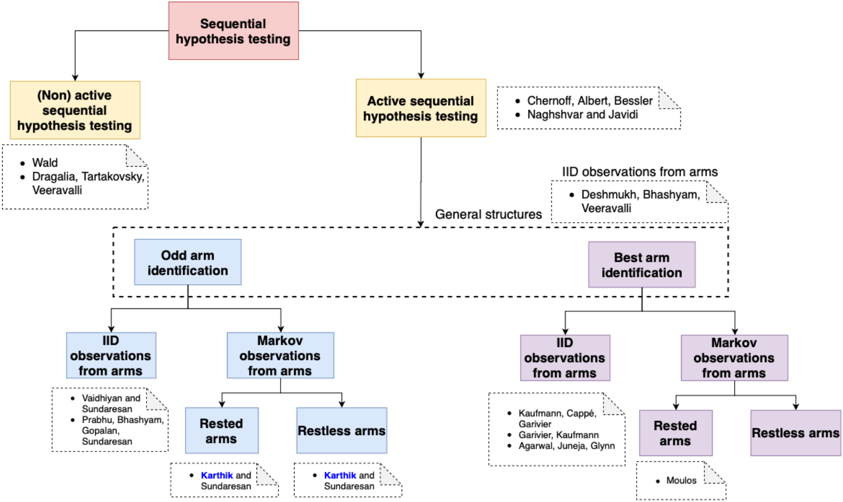
- **Fixed confidence setting**
- **Asymptotics as (b) $\rightarrow 0$.**

Two quantities of interest:

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

(b) Probability of error

The Big Picture



OAI as a Composite Hypothesis Testing Problem

When P_1 and P_2 are known: K simple hypotheses

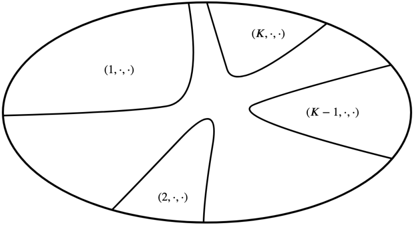
$$\begin{aligned}
\mathcal{H}_1 &: \text{arm 1 is the odd arm} \\
\mathcal{H}_2 &: \text{arm 2 is the odd arm} \\
&\vdots \\
\mathcal{H}_K &: \text{arm } K \text{ is the odd arm}
\end{aligned}$$

OAI as a Composite Hypothesis Testing Problem

When P_1 and P_2 are known: K simple hypotheses

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

When neither P_1 nor P_2 is known: K composite hypotheses



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

Notation : (odd arm, P_1, P_2)

Overview of Classical Results

Humble Beginnings

Humble Beginnings

- SPRT
- Only one arm to choose
- Optimal

- Procedure A
- Multiple arms
- Asymptotically optimal

Humble Beginnings

- SPRT
- Only one arm to choose
- Optimal

- Procedure A
- Multiple arms
- Asymptotically optimal

All's well that starts well!

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$
- To test

$$\mathcal{H}_1 : \ \theta = \theta_1$$

$$\text{vs}$$

$$\mathcal{H}_2 : \ \theta = \theta_2$$

where P_{θ_1} and P_{θ_2} are **known**

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$
- To test

$$\mathcal{H}_1 : \; \theta = \theta_1$$

$$\text{vs}$$

$$\mathcal{H}_2 : \; \theta = \theta_2$$

  where P_{θ_1} and P_{θ_2} are **known**
- Priors: $P(\mathcal{H}_1) = w = 1 - P(\mathcal{H}_2)$

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$
- To test

$$\mathcal{H}_1 : \; \theta = \theta_1$$

$$\text{vs}$$

$$\mathcal{H}_2 : \; \theta = \theta_2$$

where P_{θ_1} and P_{θ_2} are **known**
- Priors: $P(\mathcal{H}_1) = w = 1 - P(\mathcal{H}_2)$
- At each time n:
    - Stop and declare the true hypothesis
    - Take one more sample and continue

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$
- To test

$$\begin{aligned} \mathcal{H}_1 :\ & \theta = \theta_1 \\ & \text{vs} \\ \mathcal{H}_2 :\ & \theta = \theta_2 \end{aligned}$$

  where P_{θ_1} and P_{θ_2} are **<u>known</u>**
- Priors: $P(\mathcal{H}_1) = w = 1 - P(\mathcal{H}_2)$
- At each time n:
  - Stop and declare the true hypothesis
  - Take one more sample and continue
- Cost c per sample

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$
- To test

$$\mathcal{H}_1 : \; \theta = \theta_1$$
$$\text{vs}$$
$$\mathcal{H}_2 : \; \theta = \theta_2$$

  where P_{θ_1} and P_{θ_2} are **known**
- Priors: $P(\mathcal{H}_1) = w = 1 - P(\mathcal{H}_2)$
- At each time n:
  - Stop and declare the true hypothesis
  - Take one more sample and continue
- Cost c per sample
- Average risks at the stopping time N:

$$R_1 = P_{FA} + c\,\mathbb{E}[N|\mathcal{H}_1], \quad R_2 = P_{MD} + c\,\mathbb{E}[N|\mathcal{H}_2]$$

Wald's SPRT: Problem Setup

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \theta_2\}$
- To test

$$\mathcal{H}_1 : \ \theta = \theta_1$$

$$\text{vs}$$

$$\mathcal{H}_2 : \ \theta = \theta_2$$

  where P_{θ_1} and P_{θ_2} are **known**
- Priors: $P(\mathcal{H}_1) = w = 1 - P(\mathcal{H}_2)$
- At each time n:
  - Stop and declare the true hypothesis
  - Take one more sample and continue
- Cost c per sample
- Average risks at the stopping time N:

$$R_1 = P_{FA} + c\,\mathbb{E}[N|\mathcal{H}_1], \quad R_2 = P_{MD} + c\,\mathbb{E}[N|\mathcal{H}_2]$$

- Goal: minimise $w\,R_1 + (1-w)\,R_2$

Wald's SPRT: Main Idea

At time n, construct the statistic

$$S_n = \log \frac{P_{\theta_2}(X_1, \ldots, X_n)}{P_{\theta_1}(X_1, \ldots, X_n)}$$

Wald's SPRT: Main Idea

At time n, construct the statistic

$$S_n = \log \frac{P_{\theta_2}(X_1, \ldots, X_n)}{P_{\theta_1}(X_1, \ldots, X_n)}$$

We have

$$\mathbb{E}[S_n|\mathcal{H}_1] = -n\,D(P_{\theta_1}\|P_{\theta_2}), \quad \mathbb{E}[S_n|\mathcal{H}_2] = n\,D(P_{\theta_2}\|P_{\theta_1})$$

Wald's SPRT: Main Idea

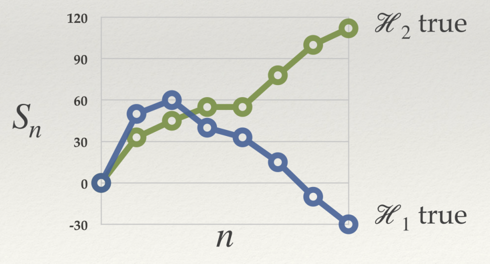
At time n, construct the statistic

$$S_n = \log \frac{P_{\theta_2}(X_1, \ldots, X_n)}{P_{\theta_1}(X_1, \ldots, X_n)}$$

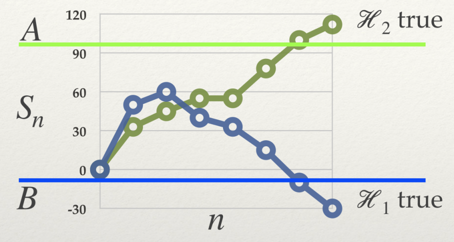
We have

$$\mathbb{E}[S_n|\mathcal{H}_1] = -n\, D(P_{\theta_1} \| P_{\theta_2}), \quad \mathbb{E}[S_n|\mathcal{H}_2] = n\, D(P_{\theta_2} \| P_{\theta_1})$$

By the law of large numbers, we expect the following behaviour:



Wald's SPRT: Algorithm



At time n:

If $S_n \geq A$, stop and declare $\mathcal{H}_2$ true

If $S_n \leq B$, stop and declare $\mathcal{H}_1$ true

If $B < S_n < A$, take one more observation

Wald's SPRT: Result

For small values of c (cost per observation):

$$A \approx -\log c, \quad B \approx \log c$$

Wald's SPRT: Result

For small values of c (cost per observation):

$$A \approx -\log c, \quad B \approx \log c$$

By Wald's identity, if SPRT stops at a random time N, then

$$\mathbb{E}[S_N|\mathcal{H}_1] = -\mathbb{E}[N|\mathcal{H}_1] \cdot D(P_{\theta_1} \| P_{\theta_2}), \quad \mathbb{E}[S_N|\mathcal{H}_2] = \mathbb{E}[N|\mathcal{H}_2] \cdot D(P_{\theta_2} \| P_{\theta_1})$$

Wald's SPRT: Result

For small values of c (cost per observation):

$$A \approx -\log c, \quad B \approx \log c$$

By Wald's identity, if SPRT stops at a random time N, then

$$\mathbb{E}[S_N|\mathcal{H}_1] = -\mathbb{E}[N|\mathcal{H}_1] \cdot D(P_{\theta_1} \| P_{\theta_2}), \quad \mathbb{E}[S_N|\mathcal{H}_2] = \mathbb{E}[N|\mathcal{H}_2] \cdot D(P_{\theta_2} \| P_{\theta_1})$$

Combining, we get

$$\boxed{\frac{\mathbb{E}[N|\mathcal{H}_1]}{\log \frac{1}{c}} \approx \frac{1}{D(P_{\theta_1} \| P_{\theta_2})}, \quad \frac{\mathbb{E}[N|\mathcal{H}_2]}{\log \frac{1}{c}} \approx \frac{1}{D(P_{\theta_2} \| P_{\theta_1})}}$$

Handling Multiple Hypotheses

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \ldots, \theta_M\}$, $M > 2$

Handling Multiple Hypotheses

- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta$, $\theta \in \{\theta_1, \ldots, \theta_M\}$, $M > 2$
- To test

$$\begin{aligned} \mathcal{H}_1 &: \ \theta = \theta_1 \\ &\vdots \\ \mathcal{H}_M &: \ \theta = \theta_M \end{aligned}$$

where $P_{\theta_1}, \ldots, P_{\theta_M}$ are **known**

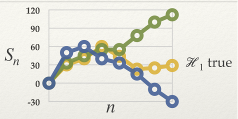
Handling Multiple Hypotheses

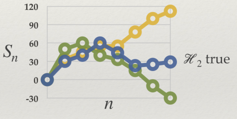
- $X_1, X_2, \ldots \overset{IID}{\sim} P_\theta,\ \theta \in \{\theta_1, \ldots, \theta_M\},\ M > 2$
- To test

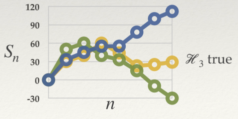
$$
\begin{aligned}
\mathcal{H}_1 &: \ \theta = \theta_1 \\
&\ \vdots \\
\mathcal{H}_M &: \ \theta = \theta_M
\end{aligned}
$$

where $P_{\theta_1}, \ldots, P_{\theta_M}$ are **known**

A behaviour we would like S_n to have (shown for $M = 3$):







M-SPRT†

Construct the statistic

$$S_n^{(i)} = \log \frac{P_{\theta_i}(X_1, \ldots, X_n)}{\max\limits_{j \neq i} P_{\theta_j}(X_1, \ldots, X_n)}, \quad i = 1, \ldots, M$$

M-SPRT[†]

Construct the statistic

$$S_n^{(i)} = \log \frac{P_{\theta_i}(X_1, \dots, X_n)}{\max\limits_{j \neq i} P_{\theta_j}(X_1, \dots, X_n)}, \quad i = 1, \dots, M$$

At time n, do the following:

Let $i^*(n) \in \arg\max\limits_i S_n^{(i)}$
If $S_n^{(i^*(n))} \geq -logc$, stop and declare $\mathcal{H}_{i^*(n)}$ true
If $S_n^{(i^*(n))} < -\log c$, take one more observation

M-SPRT†

Construct the statistic

$$S_n^{(i)} = \log \frac{P_{\theta_i}(X_1, \ldots, X_n)}{\max_{j \neq i} P_{\theta_j}(X_1, \ldots, X_n)}, \quad i = 1, \ldots, M$$

At time n, do the following:

Let $i^*(n) \in \arg\max_i S_n^{(i)}$

If $S_n^{(i^*(n))} \geq -logc$, stop and declare $\mathcal{H}_{i^*(n)}$ true

If $S_n^{(i^*(n))} < -\log c$, take one more observation

It can be shown that if N is the random stopping time of M-SPRT, then

$$\frac{\mathbb{E}[N|\mathcal{H}_i]}{\log \frac{1}{c}} \approx \frac{1}{\min_{j \neq i} D(P_{\theta_i} \| P_{\theta_j})}, \quad i = 1, \ldots, M$$

† Draglia, V. P., Alexander G. Tartakovsky, and Venugopal V. Veeravalli, "Multihypothesis sequential probability ratio tests - Part I: Asymptotic optimality", IEEE Transactions on Information Theory 45.7 (1999): 2448-2461.

Active Sequential Hypothesis Testing

- $\theta \in \{\theta_1, \ldots, \theta_M\}$, $M > 2$
- L experiments $E_1, \ldots, E_L$

Active Sequential Hypothesis Testing

- $\theta \in \{\theta_1, \ldots, \theta_M\}$, $M > 2$
- L experiments $E_1, \ldots, E_L$
- $X_1, X_2, \ldots \overset{IID}{\sim} P_{\theta_i, E_l}$ under hypothesis $\mathcal{H}_i$ and experiment E_l

Active Sequential Hypothesis Testing

- $\theta \in \{\theta_1, \ldots, \theta_M\}$, $M > 2$
- L experiments $E_1, \ldots, E_L$
- $X_1, X_2, \ldots \overset{IID}{\sim} P_{\theta_i, E_l}$ under hypothesis $\mathcal{H}_i$ and experiment E_l
- All the distributions are known

Active Sequential Hypothesis Testing

- $\theta \in \{\theta_1, \ldots, \theta_M\}$, $M > 2$
- L experiments $E_1, \ldots, E_L$
- $X_1, X_2, \ldots \overset{IID}{\sim} P_{\theta_i, E_l}$ under hypothesis $\mathcal{H}_i$ and experiment E_l
- All the distributions are known
- At time n:
    - Stop and declare the true hypothesis
    - Choose an experiment and take one more observation

Chernoff's Procedure A and its Asymptotic Optimality

Algorithm described by Chernoff:

$$S_n^{(i)} = \log \frac{P_{\theta_i}(X_1, E^{(1)}, \ldots, X_n, E^{(n)})}{\max\limits_{j \neq i} \; P_{\theta_j}(X_1, E^{(1)}, \ldots, X_n, E^{(n)})}$$

Let $i^*(n) \in \arg\max\limits_i S_n^{(i)}$
If $S_n^{(i^*(n))} \geq -logc$, stop and declare $\mathcal{H}_{i^*(n)}$ true
If $S_n^{(i^*(n))} < -\log c$, take one more observation

Chernoff's Procedure A and its Asymptotic Optimality

Algorithm described by Chernoff:

$$S_n^{(i)} = \log \frac{P_{\theta_i}(X_1, E^{(1)}, \ldots, X_n, E^{(n)})}{\max\limits_{j \neq i} \; P_{\theta_j}(X_1, E^{(1)}, \ldots, X_n, E^{(n)})}$$

Let $i^*(n) \in \arg\max\limits_i S_n^{(i)}$

If $S_n^{(i^*(n))} \geq -logc$, stop and declare $\mathcal{H}_{i^*(n)}$ true

If $S_n^{(i^*(n))} < -\log c$, take one more observation

Chernoff showed that

$$\frac{\mathbb{E}[N|\mathcal{H}_i]}{\log \frac{1}{c}} \approx \frac{1}{\max\limits_{\lambda} \min\limits_{j \neq i} \sum\limits_{l=1}^{L} \lambda(E_l)\, D(P_{\theta_i, E_l} \| P_{\theta_j, E_l})}, \quad i = 1, \ldots, M$$

Back to Odd Arm Identification

Visual Search with Static Images

- Solved by Vaidhiyan and Sundaresan
- Experiments $\equiv$ image locations
- $c \to 0 \equiv$ prob. of error $\to 0$
- Observation: number of neuron firings in the brain corresponding to an image observed
- Model: no. of firings is Poisson[†]
- Odd arm: Poisson(R_1), non-odd arms: Poisson(R_2)

† A. P. Sripati and C. R. Olson, "Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency," J. Neurosci., vol. 30, no. 4, pp. 1258–1269, Jan. 2010.

Visual Search with Static Images

When R_1 and R_2 are known[1] (prob. of error: ϵ)

$$\frac{\mathbb{E}[N|\mathcal{H}_i]}{\log \frac{1}{\epsilon}} \approx \frac{1}{\max\limits_{\lambda} \min\limits_{j \neq i} \sum_{a=1}^{K} \lambda(a)\, D(\mathrm{Poi}(R_{i,a}) \| \mathrm{Poi}(R_{j,a}))}, \quad i = 1, \ldots, M$$

[1] Vaidhiyan, Nidhin Koshy, Sripati P. Arun, and Rajesh Sundaresan. "Neural dissimilarity indices that predict oddball detection in behaviour." IEEE Transactions on Information Theory 63.8 (2017): 4778-4796.

[2] Vaidhiyan, Nidhin Koshy, and Rajesh Sundaresan. "Learning to detect an oddball target." IEEE Transactions on Information Theory 64.2 (2017): 831-852.

Visual Search with Static Images

When R_1 and R_2 are known[1] (prob. of error: ϵ)

$$\frac{\mathbb{E}[N|\mathcal{H}_i]}{\log \frac{1}{\epsilon}} \approx \frac{1}{\max_{\lambda} \min_{j \neq i} \sum_{a=1}^{K} \lambda(a)\, D(\mathrm{Poi}(R_{i,a}) \| \mathrm{Poi}(R_{j,a}))}, \quad i = 1, \ldots, M$$

When neither R_1 nor R_2 is known[2]:

$$\frac{\mathbb{E}[N|\mathcal{H}_i]}{\log \frac{1}{\epsilon}} \approx \frac{1}{\max_{\lambda} \min_{j \neq i, R_1', R_2'} \sum_{a=1}^{K} \lambda(a)\, D(\mathrm{Poi}(R_{i,a}) \| \mathrm{Poi}(R_{j,a}))}, \quad i = 1, \ldots, M$$

A policy similar to Chernoff's with forced exploration works

[1] Vaidhiyan, Nidhin Koshy, Sripati P. Arun, and Rajesh Sundaresan. "Neural dissimilarity indices that predict oddball detection in behaviour." IEEE Transactions on Information Theory 63.8 (2017): 4778-4796.

[2] Vaidhiyan, Nidhin Koshy, and Rajesh Sundaresan. "Learning to detect an oddball target." IEEE Transactions on Information Theory 64.2 (2017): 831-852.

Notion of Closeness Between Two Static Images

- The constant in the denominator captures the hardness of the problem at hand, and serves as a measure of closeness between the odd and the non-odd images
- Call this constant $D^*(i, R_1, R_2)$

Notion of Closeness Between Two Static Images

- The constant in the denominator captures the hardness of the problem at hand, and serves as a measure of closeness between the odd and the non-odd images
- Call this constant $D^*(i, R_1, R_2)$
- For two images $\mathcal{I}_1$ (odd) and $\mathcal{I}_2$ (non-odd) whose Poisson firing rates R_1 and R_2 are **known**,

$$D^*(i, R_1, R_2) = \max_{\lambda} \min_{j \neq i} \sum_{a=1}^{K} \lambda(a)\, D(\mathrm{Poi}(R_{i,a}) \| \mathrm{Poi}(R_{j,a}))$$

Notion of Closeness Between Two Static Images

- The constant in the denominator captures the hardness of the problem at hand, and serves as a measure of closeness between the odd and the non-odd images
- Call this constant $D^*(i, R_1, R_2)$
- For two images $\mathcal{I}_1$ (odd) and $\mathcal{I}_2$ (non-odd) whose Poisson firing rates R_1 and R_2 are **known**,

$$D^*(i, R_1, R_2) = \max_{\lambda} \min_{j \neq i} \sum_{a=1}^{K} \lambda(a)\, D(\mathrm{Poi}(R_{i,a}) \| \mathrm{Poi}(R_{j,a}))$$

- For two images $\mathcal{I}_1$ (odd) and $\mathcal{I}_2$ (non-odd) whose Poisson firing rates R_1 and R_2 are **not known**,

$$D^*(i, R_1, R_2) = \max_{\lambda} \min_{j \neq i, R_1', R_2'} \sum_{a=1}^{K} \lambda(a)\, D(\mathrm{Poi}(R_{i,a}) \| \mathrm{Poi}(R_{j,a}))$$

From Static Images to Movies

The Restless Odd Markov Arm Problem

- A movie is a sequence of static images (frames), each of which bears a dependency on its previous one

The Restless Odd Markov Arm Problem

- A movie is a sequence of static images (frames), each of which bears a dependency on its previous one
- When a movie is not observed, its frames move ahead

The Restless Odd Markov Arm Problem

- A movie is a sequence of static images (frames), each of which bears a dependency on its previous one
- When a movie is not observed, its frames move ahead
- Restless arms + Markov observations from each arm + one odd arm

The Restless Odd Markov Arm Problem

- A movie is a sequence of static images (frames), each of which bears a dependency on its previous one
- When a movie is not observed, its frames move ahead
- Restless arms + Markov observations from each arm + one odd arm
- This setting poses technical challenges

The Restless Odd Markov Arm Problem

- A movie is a sequence of static images (frames), each of which bears a dependency on its previous one
- When a movie is not observed, its frames move ahead
- Restless arms + Markov observations from each arm + one odd arm
- This setting poses technical challenges
- As a first step, study the setting of rested arms (movie paused when not observed)

The Restless Odd Markov Arm Problem

- A movie is a sequence of static images (frames), each of which bears a dependency on its previous one
- When a movie is not observed, its frames move ahead
- Restless arms + Markov observations from each arm + one odd arm
- This setting poses technical challenges
- As a first step, study the setting of rested arms (movie paused when not observed)

Rested Markov Arms: Setup

- $K \geq 3$ arms

Rested Markov Arms: Setup

- $K \geq 3$ arms
- Each arm is a homogeneous and ergodic DTMC on a finite state space

Rested Markov Arms: Setup

- $K \geq 3$ arms
- Each arm is a homogeneous and ergodic DTMC on a finite state space
- State space is common to all the arms

Rested Markov Arms: Setup

- $K \geq 3$ arms
- Each arm is a homogeneous and ergodic DTMC on a finite state space
- State space is common to all the arms
- Odd arm has TPM P_1, non-odd arms have TPM P_2

Rested Markov Arms: Setup

- $K \geq 3$ arms
- Each arm is a homogeneous and ergodic DTMC on a finite state space
- State space is common to all the arms
- Odd arm has TPM P_1, non-odd arms have TPM P_2
- The Markov chain of each arm evolves only when the arm is selected; otherwise, state of the arm is frozen

Rested Markov Arms: Results[3]

When P_1 and P_2 are known:

$$D^*(i, P_1, P_2) = \max_{\lambda} \min_{j \neq i} \sum_{a=1}^{K} \lambda(a)\, D(P_{i,a} \| P_{j,a} | \mu_{i,a}), \quad i = 1, \ldots, M$$

[3] P. N. Karthik and R. Sundaresan, "Learning to detect an odd markov arm," 2019. [Online]. Available: https://arxiv.org/abs/1904.11361

Rested Markov Arms: Results[3]

When P_1 and P_2 are known:

$$D^*(i, P_1, P_2) = \max_{\lambda} \min_{j \neq i} \sum_{a=1}^{K} \lambda(a)\, D(P_{i,a} \| P_{j,a} | \mu_{i,a}), \quad i = 1, \ldots, M$$

When neither P_1 nor P_2 is known:

$$D^*(i, P_1, P_2) = \max_{\lambda} \min_{j \neq i, P_1', P_2'} \sum_{a=1}^{K} \lambda(a)\, D(P_{i,a} \| P_{j,a} | \mu_{i,a}), \quad i = 1, \ldots, M$$

A policy similar to Chernoff's with forced exploration works

[3] P. N. Karthik and R. Sundaresan, "Learning to detect an odd markov arm," 2019. [Online]. Available: https://arxiv.org/abs/1904.11361

Rested Markov Arms: Insights

- Proof of the lower bound is based on a version of data processing inequality

Rested Markov Arms: Insights

- Proof of the lower bound is based on a version of data processing inequality
- Wald's identity not applicable since the observations are Markov. A generalisation needed

Rested Markov Arms: Insights

- Proof of the lower bound is based on a version of data processing inequality
- Wald's identity not applicable since the observations are Markov. A generalisation needed
- The optimum distribution does not depend on where the movie of each arm was paused. This is because of ergodicity of each arm

Restless Markov Arms

- Each arm continue to evolve whether or not it is selected (movies not paused)

Restless Markov Arms

- Each arm continue to evolve whether or not it is selected (movies not paused)
- It is important to keep track of
    - how far back in time each arm was last observed (delay)

Restless Markov Arms

- Each arm continue to evolve whether or not it is selected (movies not paused)
- It is important to keep track of
  - how far back in time each arm was last observed (delay)
  - the last observed state of each arm

Restless Markov Arms

- Each arm continue to evolve whether or not it is selected (movies not paused)
- It is important to keep track of
  - how far back in time each arm was last observed (delay)
  - the last observed state of each arm

Suppose the arm delays and last observed states at time t are

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t)), \quad \underline{s}(t) = (s_1(t), \ldots, s_K(t))$$

Restless Markov Arms

- Each arm continue to evolve whether or not it is selected (movies not paused)
- It is important to keep track of
  - how far back in time each arm was last observed (delay)
  - the last observed state of each arm

Suppose the arm delays and last observed states at time t are

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t)), \quad \underline{s}(t) = (s_1(t), \ldots, s_K(t))$$

Suppose A_t is the arm selected at time t

Restless Markov Arms

- Each arm continue to evolve whether or not it is selected (movies not paused)
- It is important to keep track of
  - how far back in time each arm was last observed (delay)
  - the last observed state of each arm

Suppose the arm delays and last observed states at time t are

$$\underline{d}(t) = (d_1(t), \ldots, d_K(t)), \quad \underline{s}(t) = (s_1(t), \ldots, s_K(t))$$

Suppose A_t is the arm selected at time t

We have

$$(\underline{d}(t), \underline{s}(t)) \longrightarrow A_t \longrightarrow (\underline{d}(t+1), \underline{s}(t+1)) \longrightarrow A_{t+1}$$

Restless Markov Setting: Key Findings[4]

- $\{(\underline{d}(t), \underline{s}(t))\}$ is a controlled Markov chain, with $\{A_t\}$ as the sequence of controls

[4] Submitted to ISIT 2020

Restless Markov Setting: Key Findings[4]

- $\{(\underline{d}(t), \underline{s}(t))\}$ is a controlled Markov chain, with $\{A_t\}$ as the sequence of controls
- When A_t's are stationary control strategies, we have an MDP on a countable state space (HARD!)

[4] Submitted to ISIT 2020

Restless Markov Setting: Key Findings[4]

- $\{(\underline{d}(t), \underline{s}(t))\}$ is a controlled Markov chain, with $\{A_t\}$ as the sequence of controls
- When A_t's are stationary control strategies, we have an MDP on a countable state space (HARD!)
- Our objective is to characterise $D^*(i, P_1, P_2)$, which is quite non-standard in MDP literature

[4] Submitted to ISIT 2020

Restless Markov Setting: Key Findings

- When A_t is chosen according to a distribution of the form

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

  for some $\lambda(\cdot \mid \cdot)$ and $\eta > 0$, the Markov process $\{(\underline{d}(t), \underline{s}(t))\}$ becomes ergodic

Restless Markov Setting: Key Findings

- When A_t is chosen according to a distribution of the form

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

  for some $\lambda(\cdot \mid \cdot)$ and $\eta > 0$, the Markov process $\{(\underline{d}(t), \underline{s}(t))\}$ becomes ergodic

- When P_1 and P_2 are known,

$$D^*(i, P_1, P_2) = \sup_{\lambda(\cdot|\cdot)} \min_{j \neq i} \sum_{(\underline{d}, \underline{s})} \sum_{a=1}^{K} \nu^{\lambda}(\underline{d}, \underline{i}, a)\, D(P_{i,a}^{d_a}(\cdot|s_a) \| P_{j,a}^{d_a}(\cdot|s_a))$$

- Sample complexity-type results
- Generalises all the previously known results

Open Threads

- Analysis of the restless arms setting for the case when neither P_1 nor P_2 is known

Open Threads

- Analysis of the restless arms setting for the case when neither P_1 nor P_2 is known
- Characterising the second-order term:

$$\mathbb{E}[N|\mathcal{H}_i] \approx \frac{\log \frac{1}{\epsilon}}{D_i^*} + ?? \quad \left(\text{e.g., } ?? = C \cdot \log\log \frac{1}{\epsilon}\right)$$

Open Threads

- Analysis of the restless arms setting for the case when neither P_1 nor P_2 is known
- Characterising the second-order term:

$$\mathbb{E}[N|\mathcal{H}_i] \approx \frac{\log \frac{1}{\epsilon}}{D_i^*} + ?? \quad \left(\text{e.g., } ?? = C \cdot \log\log \frac{1}{\epsilon}\right)$$

- Analysis for the case when each arm is a Markov process on a general state space

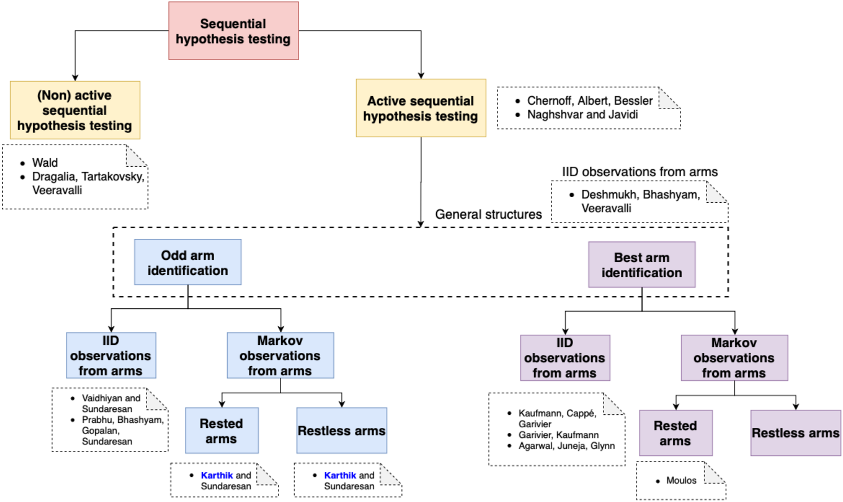
Open Threads

- Analysis of the restless arms setting for the case when neither P_1 nor P_2 is known
- Characterising the second-order term:

$$\mathbb{E}[N|\mathcal{H}_i] \approx \frac{\log \frac{1}{\epsilon}}{D_i^*} + ?? \quad \left(\text{e.g., } ?? = C \cdot \log \log \frac{1}{\epsilon} \right)$$

- Analysis for the case when each arm is a Markov process on a general state space
- General structures for Markov observations

A Final Glimpse of Where We Stand



Thank You!