# Ph. D. Thesis Synopsis
# Department of Electrical Communication Engineering
# Indian Institute of Science

July 2, 2021

Name of the Candidate: P. N. Karthik
Thesis Supervisor: Dr. Rajesh Sundaresan
Thesis Title: Sequential Controlled Sensing to Detect an Anomalous Process

In this thesis, we study the problem of identifying an anomalous arm in a multi-armed bandit as quickly as possible, subject to an upper bound on the error probability. Also known as odd arm identification, this problem falls within the class of optimal stopping problems in decision theory and can be embedded within the framework of active sequential hypothesis testing. Prior works on odd arm identification dealt with independent and identically distributed observations from each arm. We provide the first known extension to the case of Markov observations from each arm. Our analysis and results are in the asymptotic regime of vanishing error probability.

We associate with each arm an ergodic discrete time Markov process that evolves on a common, finite state space. The notion of anomaly is that the transition probability matrix (TPM) of the Markov process of one of the arms (the *odd arm*) is some $P_1$, and that of each non-odd arm is a different $P_2$, $P_2 \neq P_1$. A learning agent whose goal it is to find the odd arm samples the arms sequentially, one at any given time, and observes the state of the Markov process of the sampled arm. The Markov processes of the unobserved arms may either remain frozen at their last observed states until their next sampling instant (*rested arms*) or continue to undergo state evolution (*restless arms*). The TPMs $P_1$ and $P_2$ may be known to the learning agent beforehand or unknown.

Given an upper bound $\epsilon \in (0,1)$ on the learning agent's error probability, intuitively, the smaller the value of $\epsilon$, the longer the learning agent will have to wait before finding the odd arm. Let $\pi$ be any sequential arm selection policy of the learning agent whose time to find the odd arm is denoted by $\tau(\pi)$ (a random variable). Let $\Pi(\epsilon)$ denote the collection of all policies whose error probability is no more than $\epsilon$. Let $C = (h, P_1, P_2)$ denote the problem instance in which $h$ is the index of the odd arm, $P_1$ is the TPM of arm $h$, and $P_2 \neq P_1$ is the TPM of each non-odd arm. We anticipate from the prior works that $\inf\limits_{\pi \in \Pi(\epsilon)} E[\tau(\pi)|C] = O\left(\log \frac{1}{\epsilon}\right)$. Here, $E[\tau(\pi)|C]$ denotes the expectation of $\tau(\pi)$ computed under the problem instance $C = (h, P_1, P_2)$. Our objective is to characterise

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

The value of (1) depends, in general, on (a) $P_1$ and $P_2$, (b) whether the TPMs $P_1$ and $P_2$ are known beforehand or unknown, and (c) whether the arms are rested or restless. From the symmetry of the problem, it can be deduced that (1) does not depend on the index of the odd arm. We analyse the following cases: (a) rested arms with TPMs unknown, (b) restless arms with TPMs known, and (c) restless arms with TPMs unknown. For each of these cases, we provide an explicit answer for (1) by first presenting an asymptotic lower bound on (1) and subsequently devising a sequential arm selection policy that achieves this lower bound and is therefore asymptotically optimal.

A key ingredient in our analysis of the setting of rested arms is the observation that for the Markov process of each arm, the long term fraction of entries into a state is equal to the long term fraction of exits from the state (global balance). When the arms are restless, it is necessary for the learning agent to keep track of the time since each arm was last sampled (arm's *delay*) and the state of each arm when it was last sampled (arm's *last observed state*). We show that the arm delays and the last observed states form a controlled Markov process which is ergodic under any stationary arm selection policy that picks each arm with a strictly positive probability. Our approach of considering the delays and the last observed states of all the arms jointly offers a global perspective of the arms and serves as a 'lift' from the local perspective of dealing with the delay and the last observed state of each arm separately, one that is suggested by the prior works. Lastly, when the TPMs are unknown and have to be estimated along the way, it is important to ensure that the estimates converge almost surely to their true values asymptotically, i.e., the system is *identifiable*. We show identifiability follows from the ergodic theorem in the rested case, and provide sufficient conditions for it in the restless case.

**Publications Based on this Thesis**


1. P. N. Karthik, R. Sundaresan, *Learning to Detect an Odd Restless Markov Arm with a Trembling Hand*, submitted.
2. P. N. Karthik, R. Sundaresan, *Detecting an Odd Restless Markov Arm with a Trembling Hand*, IEEE Transactions on Information Theory, July 2021.
3. P. N. Karthik, R. Sundaresan, *Learning to Detect an Odd Markov Arm*, IEEE Transactions on Information Theory, July 2020, vol. 66, no. 7, pp. 4324 – 4348.
4. P. N. Karthik, R. Sundaresan, *Learning to Detect an Odd Restless Markov Arm*, proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), virtual conference.
5. P. N. Karthik, R. Sundaresan, *Detecting an Odd Restless Markov Arm with a Trembling Hand*, proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT), virtual conference.
6. P. N. Karthik, R. Sundaresan, *Learning to Detect an Odd Markov Arm*, proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, July 2019.