Adaptive Control and Parameter Estimation in Markov Chains: A Quadratic Case

By

Stephanie N. Walker

Submitted to the graduate degree program in Mathematics
and the Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Master of Arts.

Chair: Dr. Bozenna Pasik-Duncan

Dr. Margaret Bayer

Dr. Tyrone Duncan

Date defended: 28 April 2022

The Thesis Committee for Stephanie N. Walker

certifies that this is the approved version of the following thesis:

Adaptive Control and Parameter Estimation in Markov Chains: A Quadratic Case

Chair: Dr. Bozenna Pasik-Duncan

Date approved: 12 May 2022

Abstract

The objective of this thesis is to extend results to a new quadratic case in support of a collection of existing results in adaptive control and parameter estimation of Markov chains.

In the first chapter, we introduce Markov chains and discuss some of their important properties. This chapter will help the reader understand the characteristics of Markov chains, which will be useful in later chapters discussing results on adaptive control of these stochastic processes. In the second chapter, we introduce Martingales and discuss some of their important properties. Martingales are important tools used in the methods of parameter estimation of Markov chains in the later chapters.

In the third chapter, we focus on adaptive control of Markov chains. First, we consider controlled Markov chains and some general connections with martingales, namely the Law of Large Numbers and the Central Limit Theorem of Markov chains. Then, we introduce the adaptive control environment with a controlled Markov chain with an unknown parameter. It is after this that we discuss previous important results in [6],[1], and [3] on adaptive control and parameter estimation of Markov chains. Then in the fourth chapter, we outline the processes used and results attained in [8] for a linear case of adaptive control of Markov chains, which we then extend to a quadratic case in the final chapter.

For the main results, we perform the process discussed in [7] for the following problem. We consider a controlled Markov chain with a finite state space, whose transition probabilities are assumed to depend quadratically on an unknown real parameter α. Particularly, we study the behavior of the maximum likelihood estimate of α at each time n as n increases under an arbitrary realizable control. We show that the results of [8] extend to the quadratic case with a few additional assumptions. These results are that the sequence of estimates of α converge almost surely, though not necessarily to the true parameter. We characterize those realizations for which convergence does not lead to the true value, and suggest corrections to the control to attain convergence to the true value. In support of previous results, we show that the maximum likelihood estimate converges to a value α^* indistinguishable from the true value under a control feedback law induced by α^*.

Acknowledgements

I would like to thank my research advisor Dr. Bozenna Pasik-Duncan for providing guidance and encouragement. She was able to assess my interests and direct me through an exciting research journey.

I would also like to thank Dr. Tyrone Duncan and Dr. Margaret Bayer for serving on my Thesis Committee. Their questions and suggestions helped me fine tune my thesis and also consider new viewpoints for further investigations.

I would like to thank the Mathematics and Statistics Department at the University of Central Oklahoma and Mathematics Department at the University of Kansas for supporting me through my higher education.

I would like to specifically thank Dr. Britney Hopkins, Dr. Michael Fulkerson, and Dr. Scott Williams at the University of Central Oklahoma for inspiring me to further my education at the University of Kansas.

I would also like to give credit to the Math 750 course on Stochastic Adaptive Control for introducing me to a field of mathematics that I had not seen before.

Finally, I would like to thank my parents for supporting me through my education, especially through the times of high stress, and cheering me on to success.

Contents

1 An Overview of Markov Chains

1.1 Introduction

Markov chains are stochastic models describing sequences of possible events that satisfy the Markov property:

Property 1 (Markov Property)**.** *A stochastic process that satisfies the Markov property is "memoryless", so the probability of each event only depends on the state attained from the previous event, i.e. for a sequence of random variable* $\{X_n\}_{n\geq 0}$ *and possible states* $i_0, \ldots, i_{n+1}$*,*

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n).$$

Markov chains are very useful stochastic processes due to their "memorylessness," as all necessary information for prediction is known based on the current state. Since the collection of states can be many different things, Markov chains have been used to model things such as weather forecasts, stock prices and GDP growth, population dynamics, and various games of chance. This paper will focus on discrete-time Markov chains.

Definition 1. *A* **discrete-time Markov chain** *is a stochastic process with a finite or countable state space that satisfies the Markov property.*

The changes of state in a Markov chain are called transitions, and the probability of these changes of state are called transition probabilities.

Definition 2. *A* **homogeneous Markov chain** *is a Markov chain wherein the transition probabilities are independent of time* n*.*

Definition 3. *The* **one-step transition probability** *is the probability of going from state* i *at time* n *to state* j *at time* $n+1$*, denoted*

$$P(X_{n+1} = j | X_n = i) = P_{ij}^{n,n+1}.$$

For a homogeneous Markov chain, these transition probabilities are denoted as simply P_{ij} as they are independent of time. These one-step transition probabilities form a transition probability

matrix $\mathcal{P} = [P_{ij}]$ with the number of columns and number of rows equal to the number of states. This matrix is a stochastic matrix, so the entries $P_{ij} \geq 0$ for all i, j and $\sum_j P_{ij} = 1$ for all i. A Markov chain is completely characterized by a state space, a transition probability matrix $\mathcal{P}$, and an initial state distribution across the state space.

Markov chains can be used to predict future behavior beyond the "next step". Considering the passage of some time n, we can make predictions on a state obtained at time $m + n$ based on the current state at time m.

Definition 4. *The* ***n*-step transition probability** *is the transition probability of going from state i to state j in n steps, denoted by*

$$P(X_{m+n} = j | X_m = i) = P_{ij}^{(n)}.$$

Recursively, these transition probabilities can be found by

$$P_{ij}^{(n)} = \sum_k P_{ik} P_{kj}^{(n-1)}, \qquad P_{ij}^{(1)} = P_{ij},$$

summing over all states k. The n-step transition probability matrix $\mathcal{P}^{(n)} = \mathcal{P}^n$.

1.2 Irreducible Markov Chains

The states of a Markov chain and how they relate to one another are key characteristics in prediction by Markov processes.

Definition 5.

(i) *State j is* **accessible from state *i*** *if there is a positive probability that state j can be reached from state i in a finite number of transitions, i.e. $P_{ij}^{(n)} > 0$ for some finite time n.*

(ii) *State i and state j* **communicate** *if they are each accessible from the other. This is denoted $i \sim j$.*

Note, communication between states $i \sim j$ is an equivalence relation which yields equivalence classes, called communicating classes. Then, we can study the communicating classes of a Markov chain, and even focus specifically on the communicating class induced by our current state in order

to make state predictions for a future time. This information could tell us a few useful things. First, we could determine that a specific initial state i virtually cuts off some states as possible future states because they are not accessible from state i. Thus, our predictions would be limited to states that are accessible from i. Another notable case is defined below, in which a Markov chain contains only one communicating class.

Definition 6. *A Markov chain is* **irreducible** *if all states communicate.*

Irreducible Markov chains let us know that at any time n, we cannot eliminate any state from the foreseeable future of our process. If we have multiple communicating classes, there are at least two states that do not communicate with each other. Then, being in one of these states tells us that we will never reach the other in the future of our process.

Definition 7. *A set of states $\mathcal{C}$ is* **closed** *if no one-step transition is possible from a state in $\mathcal{C}$ to a state outside of $\mathcal{C}$.*

If there are no sets of states that are closed other than the set of all states, then the Markov chain is irreducible. Consider the following example of a transition probability matrix:

Example 1. Let $\mathcal{P}$ be a transition probability matrix of the form

$$\mathcal{P} = \begin{bmatrix} A_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & A_n \end{bmatrix},$$

where A_k are stochastic matrices. Then, we can determine that $\mathcal{P}$ and the corresponding Markov chain are not irreducible as each A_k represents a closed set of states: no one-step transition is possible from states described in A_k to states described in A_ℓ. In this case, we say that $\mathcal{P}$ is decomposable.

1.3 Classification of States

The states in a Markov chain can be further classified by the following:

Definition 8.

(i) *The* **period of state *i***, *denoted* $d(i)$, *is the greatest common divisor of all integers* $n \geq 1$ *for which there is a positive probability of going from state* i *back to state* i *in* n *steps, i.e.* $P_{ii}^{(n)} > 0$.

(ii) *A state* i *is* **periodic** *if* $d(i) \geq 2$.

(iii) *A Markov chain is* **aperiodic** *if every state has period* $d(i) = 1$.

Proposition 1. *If* $i \sim j$, *then* $d(i) = d(j)$.

By this proposition, we have that any communicating class has a constant period among all the states in the class. Thus, if we know the period of one state, we know the period of all states in that communicating class.

For the next set of definitions, we will consider the return to a state i. Let us define for $n \geq 1$ the probability that, after starting in state i, the first return to state i occurs at time n as

$$f_{ii}^{(n)} = P(X_n = i, X_m \neq i \text{ for } m = 1, \ldots, n-1 | X_0 = i).$$

From this concept of first return, we can define the probability that a process starting in state i returns to state i in some finite time as

$$f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)},$$

where $f_{ii}^{(0)} = 0$.

Definition 9.

(i) *A state* i *is* **recurrent** *if after the process begins in state* i, *the probability of returning to state* i *in a finite number of steps is 1, i.e.* $f_{ii} = 1$.

(ii) *A state* i *is* **transient** *if after the process begins in state* i, *there is a positive probability of never returning to state* i *(i.e. state* i *is nonrecurrent). For a transient state* i, $f_{ii} < 1$.

Proposition 2. *Suppose* $i \sim j$. *If* i *is recurrent, then* j *is recurrent.*

Like proposition 1, this gives us information on an entire communicating class. If any state in the communicating class is recurrent, we know that all states in the class are recurrent.

We can further classify recurrent states. To do so, let us first define the mean recurrence time m_i. Let $R_i = \min\{n \geq 1 | X_n = i\}$ be the first return time, and then define

$$m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}.$$

Definition 10.

(i) *A recurrent state i is* **null** *if the average recurrence time is infinite, i.e.* $m_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)} = \infty$.

(ii) *A recurrent state i is* **positive-recurrent** *if it is not a null state.*

(iii) *A recurrent state i is* **ergodic** *if it is neither null nor periodic.*

Note, a recurrent state in a Markov chain with finite states cannot be null as the average recurrence time will be finite. Furthermore, a Markov chain with finite states cannot contain only transient states by the Pigeonhole principle as eventually we would run out of new states to obtain at some finite time. If all states in a Markov chain are ergodic, then the chain is ergodic. Ergodic implies irreducible, and we can determine that every state will be obtained at some finite time.

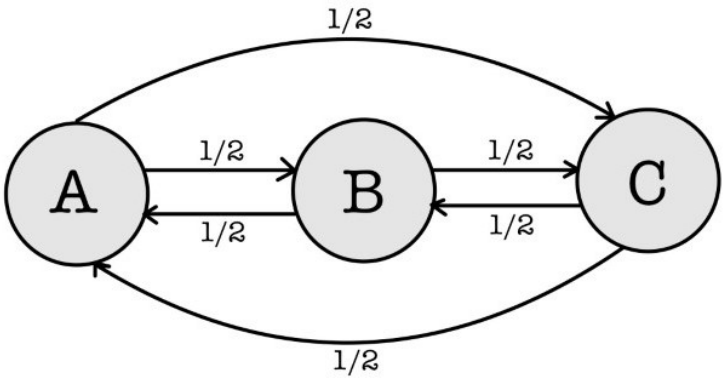
Definition 11. *A state i is* **absorbing** *if once state i is reached, it cannot be left.*

1.4 Examples of Markov Chains

Example 2. In this example, let us classify the states for a Markov chain with transition probability matrix

$$\mathcal{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

A diagram of this Markov chain is shown below:



Let us classify each state in this chain:

- State A: Notice, starting in state A, we can return to state A in 2 steps ($A \to B \to A$) and also in 3 steps ($A \to B \to C \to A$). Note, $\gcd(2,3) = 1$, so state A has period $d(A) = 1$. Also, notice A is recurrent. Therefore state A is ergodic.
- State B: Starting in state B, we can return to state B in 2 steps ($B \to A \to B$) and in 3 steps ($B \to C \to A \to B$), so state B has period $d(B) = 1$. Since state B is also recurrent, state B is ergodic.
- State C: Starting in state C, we can return to state C in 2 steps and in 3 steps, so state C has period $d(C) = 1$. Since state C is also recurrent, state C is ergodic.

Note A is accessible from B and C, B is accessible from A and C, and C is accessible from A and B. Then we know

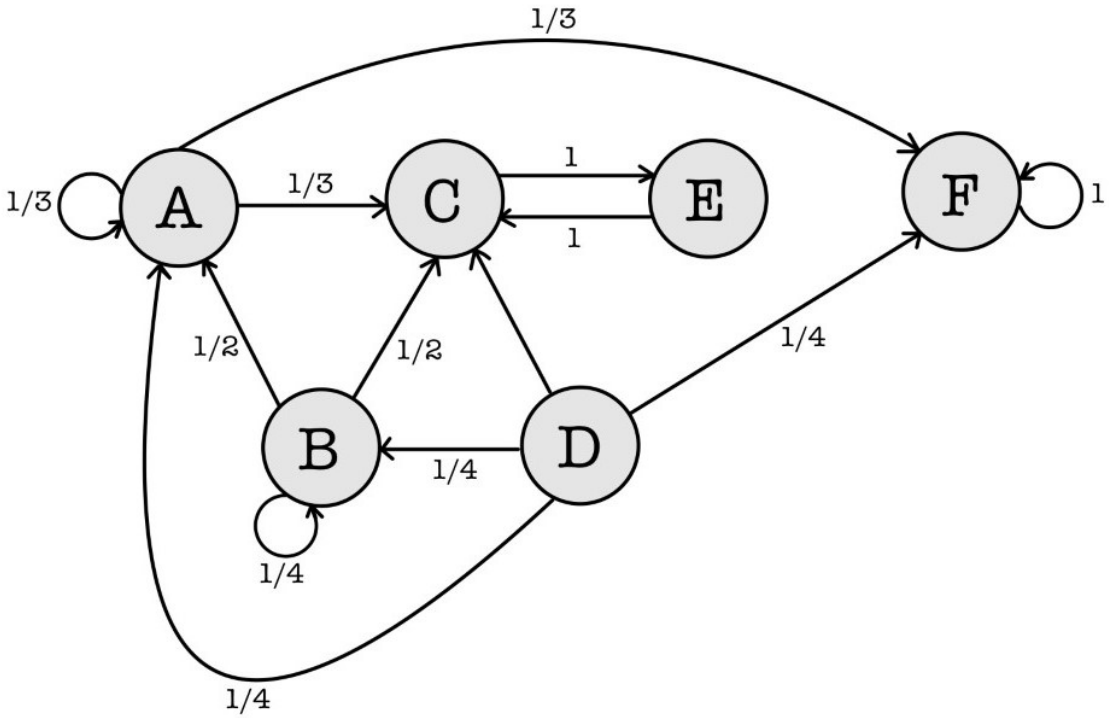
$$A \sim B, \quad A \sim C, \quad B \sim C.$$

Then alternatively, since A, B, C are in the same communicating class, after determining state A had period $d(A) = 1$ and was recurrent we could conclude that states B and C also had period 1 and were recurrent. Also, since every state communicates, our Markov chain is irreducible. Since every state has period 1, our Markov chain is aperiodic.

Example 3. In this example, let us classify the states for a Markov chain with transition probability matrix

$$\mathcal{P} = \begin{bmatrix} 1/3 & 0 & 1/3 & 0 & 0 & 1/3 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

A diagram of this Markov chain is shown below:



Classification of states:

- State A has period $d(A) = 1$ as we have the possible path $A \to A$, and state A is transient as the paths through state C and state F will never return to A.
- State B has period $d(B) = 1$ and is transient as the paths through state A and state C with never return to B.
- State C has period $d(C) = 2$ as the only paths from state C to state C are of the form $C \to E \to C \to \ldots \to E \to C$. State C is recurrent.

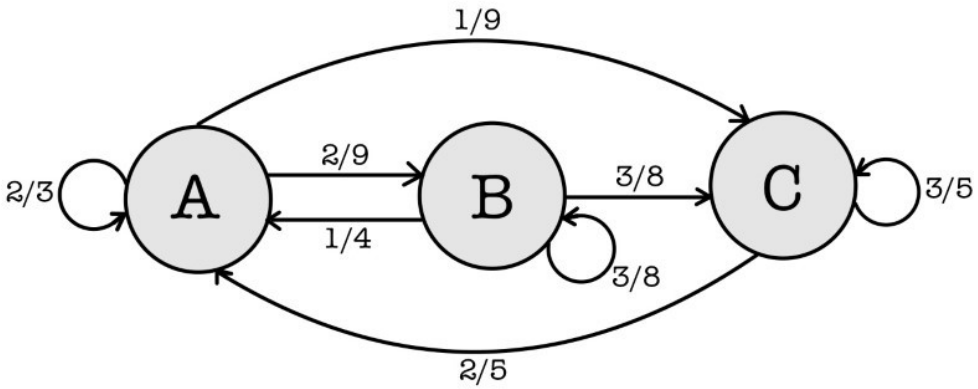
- State D has period $d(D) = 0$ and is transient as no path from state D will ever return.
- State E has period $d(E) = 2$ and is recurrent as $E \sim C$.
- State F has period $d(F) = 1$ and is recurrent, so state F is ergodic. Notice that state F is also absorbing as nothing can leave state F.

Lastly, notice $\{C, E\}$ form a closed set as after state C is attained, only states E and C are attainable in the future.

Example 4. For this example, let us create a model for fall weather in Lawrence, Kansas. Define our state space as $\{A = \text{“sunny”}, B = \text{“cloudy”}, C = \text{“stormy”}\}$. Say given a sunny day, the probability the next day is also sunny is $\frac{2}{3}$, the probability the next day is cloudy is $\frac{2}{9}$, and the probability the next day is stormy is $\frac{1}{9}$. On a cloudy day, say the probability the next day is sunny is $\frac{1}{4}$, the probability the next day is cloudy is $\frac{3}{8}$, and the probability the next day is stormy is $\frac{3}{8}$. On a stormy day, say the probability the next day is sunny is $\frac{2}{5}$ and the probability the next day is stormy is $\frac{3}{5}$. Now we have the transition probability matrix

$$\mathcal{P} = \begin{bmatrix} 2/3 & 2/9 & 1/9 \\ 1/4 & 3/8 & 3/8 \\ 2/5 & 0 & 3/5 \end{bmatrix}.$$

A diagram of this Markov chain is shown below:



Classification of states:

- State A has period $d(A) = 1$ and is recurrent. Therefore state A is ergodic.

- State B has period $d(B) = 1$ and is recurrent as $A \sim B$. Therefore state B is ergodic.
- State C has period $d(C) = 1$ and is recurrent as $A \sim C$. Therefore state C is ergodic.

Note, since every state communicates, our Markov chain is irreducible. Since every state has period 1, our Markov chain is aperiodic.

These classifications make sense considering the natural weather transitions we can observe: sunny, cloudy, and stormy weather are recurrent, can occur multiple days in a row (period 1), and each state sunny, cloudy, and stormy is accessible from every other state.

1.5 Long Run Behavior of Markov Chains

An important consideration of any system is its long run behavior. In this section, we will consider the long run behavior of Markov chains.

Definition 12. *Let $\mathcal{P}$ be a transition probability matrix on a finite number of states $1, \ldots, N$. $\mathcal{P}$ is* **regular** *if there exists a positive integer k such that $\mathcal{P}^k$ has all positive entries, i.e. $P_{ij}^{(k)} > 0$ for all i, j.*

A Markov chain with a regular transition probability matrix is also said to be regular. Regular Markov chains possess a limiting probability distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$, where $\pi_j > 0$ for all j and $\sum_j \pi_j = 1$:

$$\lim_{n\to\infty} P(X_n = j | X_0 = i) = \pi_j > 0$$

for all $j = 1, \ldots, N$. Note, this distribution is independent of the initial state. Then we can conclude that in the long run, the probability of the Markov chain being in state j is almost surely π_j, independent of our initial state at time 0.

We can find the limiting distribution π_j from the system of N linear equations

$$\pi_j = \sum_{i=1}^{N} \pi_i P_{ij}$$

for $j = 1, \ldots, N$ where $\sum_j \pi_j = 1$.

Let us denote the matrix of limiting distributions $(\pi_1, \ldots, \pi_N)$ corresponding to a regular

Markov chain as

$$\Pi = \begin{bmatrix} \pi_1 & \cdots & \pi_N \\ \vdots & & \vdots \\ \pi_1 & \cdots & \pi_N \end{bmatrix}.$$

Proposition 3. *For a regular transition probability matrix $\mathcal{P}$ and limiting distributions matrix Π, we have*

$$\lim_{n\to\infty} \mathcal{P}^n = \Pi.$$

Define matrix T such that for transition probability matrix $\mathcal{P}$ and limiting distributions matrix Π

$$\mathcal{P} = \Pi + T.$$

Proposition 4. *For transition probability matrix $\mathcal{P}$, limiting distributions matrix Π, and matrix T s.t. $\mathcal{P} = \Pi + T$,*

$$\mathcal{P}^n = \Pi + T^n.$$

Example 5. Consider state space $\{1, 2\}$ and transition probability matrix

$$\mathcal{P} = \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix}.$$

Note $\mathcal{P}^{(1)}$ has all positive entries, so $\mathcal{P}$ is regular. Let us find the limiting distributions. We have the system

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_1 + \frac{2}{5}\pi_2 \\ \pi_2 = \frac{1}{2}\pi_1 + \frac{3}{5}\pi_2 \end{cases}$$

Then, since $\pi_1 + \pi_2 = 1$, we have $\pi_1 = 1 - \pi_2$. Notice by substituting, we have

$$\begin{aligned}
\pi_1 = 1 - \pi_2 &\implies \pi_2 = \frac{1}{2}(1 - \pi_2) + \frac{3}{5}\pi_2 \\
&\implies \pi_2 = \frac{1}{2} - \frac{1}{2}\pi_2 + \frac{3}{5}\pi_2 \\
&\implies \frac{3}{2}\pi_2 - \frac{3}{5}\pi_2 = \frac{1}{2} \\
&\implies \frac{9}{10}\pi_2 = \frac{1}{2} \\
&\implies \pi_2 = \frac{5}{9}.
\end{aligned}$$

Then $\pi_2 = \dfrac{5}{9} \implies \pi_1 = 1 - \dfrac{5}{9} = \dfrac{4}{9}$. Thus, in the long run we have the probability of this Markov chain being in state 1 is $4/9$, and the probability of this Markov chain being in state 2 is $5/9$.

1.6 Markov Chains with Returns

Now consider a Markov chain with state space $\{1, \ldots, N\}$ in which a return is gained for transitioning from state i to state j. Denote this return r_{ij}, and let us define the return matrix

$$R = [r_{ij}]$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, N$.

Now, just as we were interested in predicting what state a Markov chain would have attained after some time n, if our Markov chain yields returns after each transition we might also be interested in predicting the accumulated returns after time n. Let $v_i(n)$ be the expected return after n-steps, assuming the Markov chain is in state i at time $t = 0$. Let

$$q_i = \sum_{j=1}^{N} p_{ij} \cdot r_{ij},$$

which represents the expected return after one step. Then for $i = 1, \ldots, N$,

$$v_i(n) = \sum_{j=1}^{N} \Big[p_{ij}\big(r_{ij} + v_j(n-1)\big) \Big] = q_i + \sum_{j=1}^{N} p_i j \cdot v_j(n-1),$$

and $V_i(0) = 0$. In vector form, we have $\mathbf{v}(n) = \mathbf{q} + \mathcal{P}\mathbf{v}(n-1)$.

Proposition 5. *For probability transition matrix $\mathcal{P}$, limiting distributions matrix Π, and matrix T s.t. $\mathcal{P} = \Pi + T$, we have that*

$$\mathbf{v}(n) = n\Pi\mathbf{q} + (I - T)^{-1}(I - T^n)\mathbf{q}.$$

Note, by propositions 3 and 4,

$$\lim_{n\to\infty} \mathcal{P}^n = \Pi \implies \lim_{n\to\infty} (\Pi + T^n) = \Pi \implies \lim_{n\to\infty} T^n = 0.$$

Then for large n we have

$$\mathbf{v}(n) = n\Pi\mathbf{q} + \mathbf{v}$$

where $\mathbf{v} = (I - T)^{-1}\mathbf{q}$. Now, notice

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{v}(n) = \lim_{n\to\infty} \frac{1}{n}\left(n\Pi\mathbf{q} + \mathbf{v}\right) = \lim_{n\to\infty} \left(\Pi\mathbf{q} + \frac{1}{n}\mathbf{v}\right) = \Pi\mathbf{q}.$$

Let $\mathbf{g} = \Pi\mathbf{q}$. If $\mathcal{P}$ is regular, then the rows of Π are the same, and $g_i = g_j$ for all $i, j \in \{1, \ldots, N\}$. Thus, let us define

$$g = \sum_{i=1}^{N} \pi_i g_i.$$

We say that g is the value of the game on the Markov chain with transition probability matrix $\mathcal{P}$ and return matrix R.

2 An Overview of Martingales

2.1 Introduction

Martingales are stochastic processes in which the conditional expectation of the next value in the process is the current value, regardless of past values.

Definition 13. *A* **discrete-time martingale** *is a stochastic process $\{X_n\}_{n\geq 0}$ in which for any time n,*

$$E(X_{n+1}|X_0, X_1, \ldots, X_n) = X_n.$$

Similarly to Markov chains, martingales are useful stochastic processes as information about the past is not useful for predicting the future of the process. For this reason, martingales are used to model "fair" games as ones future success in the game does not depend on past success or failures. Discrete-time martingales are also used in biodiversity and biogeography to model the number of individuals of a particular species of a fixed size at any given time.

2.2 Examples of Martingales

Example 6. An unbiased random walk, a stochastic process that describes a path of successive random steps on some space, is a martingale. For a simple example, consider a walk on $\mathbb{Z}$ starting at 0, where each step either contributes $+1$ or -1 to the position. For this walk to be unbiased, each step has an equal probability (p = 0.5) to move $+1$ or -1. Say at time n, our position $X_n = a$. Then notice

$$E(X_{n+1}|X_n = a) = 0.5(a+1) + 0.5(a-1) = 0.5a + 0.5 + 0.5a - 0.5 = a = X_n.$$

Thus, our unbiased random walk is a martingale.

Example 7. A gambler's fortune is a martingale if all the betting games the gambler plays are fair. Let X_n be a gambler's fortune after n tosses of a fair coin, i.e. $p = 0.5$ a toss comes up heads, $p = 0.5$ a toss comes up tails. Suppose the gambler wins \$5 if the coin comes up heads and loses \$5 if it comes up tails. Then, notice

$$E(X_{n+1}|X_n) = 0.5(X_n + 5) + 0.5(X_n - 5) = 0.5X_n + 0.25 + 0.5X_n - 0.25 = X_n.$$

Thus the gambler's fortune is a martingale for fair betting games. As in example 6, the symmetry of the "unbiased" and "fair" assumptions yields our desired martingale result.

Example 8. Pólya's urn is a statistical model in which we have an urn containing a number of different colored marbles, and each iteration involves drawing a random marble, noting the color, and returning the marble along with k more marbles of the same color. For a given color, the fraction of marbles of said color in the urn is a martingale.

Suppose on the n-th iteration, there are 75 green marbles and 25 non-green marbles in the

urn. Let $k = 5$, so whichever color marble is drawn next, 5 more marbles of that same color will be added to the urn. So, we have $X_n = 75/100 = 0.75$ in respect to he number of green marbles, and

$$\begin{aligned} E(X_{n+1}|X_n = 75/100) &= (75/100)(80/105) + (25/100)(75/105) \\ &= 60/105 + 74/420 \\ &= 0.75 \\ &= X_n. \end{aligned}$$

Thus, the fraction of green marbles in the urn is a martingale.

2.3 Submartingales and Supermartingales

A few generalizations of the martingale concept are submartingales and supermartingales. In these cases, future prediction still does not depend on the past of the process, however the conditional expectation of the next value is not necessarily the current value. Rather, the current value gives an upper or lower bound of the expected next value.

Definition 14. *A* **discrete-time submartingale** *is a stochastic process* $\{X_n\}_{n\geq 0}$ *in which for any time* n*,*

$$E(X_{n+1}|X_0, X_1, \ldots, X_n) \geq X_n.$$

Definition 15. *A* **discrete-time supermartingale** *is a stochastic process* $\{X_n\}_{n\geq 0}$ *in which for any time* n*,*

$$E(X_{n+1}|X_0, X_1, \ldots, X_n) \leq X_n.$$

Following are a few notable propositions pertaining to submartingales and supermartingales:

Proposition 6. *If* $\{X_n\}_{n\geq 0}$ *is a submartingale, then* $\{-X_n\}_{n\geq 0}$ *is a supermartingale. Similarly, if* $\{X_n\}_{n\geq 0}$ *is a supermartingale, then* $\{-X_n\}_{n\geq 0}$ *is a submartingale.*

Proposition 7.

(i) *If* $\{X_n\}_{n\geq 0}$ *is a martingale, then the sequence* $\{E(X_n)\}_{n\geq 0}$ *is constant.*

(ii) *If* $\{X_n\}_{n\geq 0}$ *is a submartingale, then the sequence* $\{E(X_n)\}_{n\geq 0}$ *is increasing.*

(iii) *If* $\{X_n\}_{n\geq 0}$ *is a supermartingale, then the sequence* $\{E(X_n)\}_{n\geq 0}$ *is increasing.*

Proposition 8. *Suppose* $\{X_n\}_{n\geq 0}$ *is a martingale.*

(i) *If* f *is a convex function, then* $\{f(X_n)\}_{n\geq 0}$ *is a submartingale.*

(ii) *If* f *is a concave function, then* $\{f(X_n)\}_{n\geq 0}$ *is a supermartingale.*

Particularly, if $\{X_n\}_{n\geq 0}$ is a martingale, $\{|X_n|\}_{n\geq 0}$ and $\{X_n^2\}_{n\geq 0}$ are positive submartingales. The inequalities in the next section utilize these submartingales.

Proposition 9. *Suppose* $\{X_n\}_{n\geq 0}$ *is a submartingale. Then for all* n*,*

$$E(X_{n+1}) \geq E(X_n).$$

2.4 Martingale Convergence Theorem

Among all stochastic processes, an important consideration is whether a process will converge. This gives strong prediction information for the future of the process. Before presenting the Martingale Convergence Theorem, we will define different types of convergence of random variables.

Definition 16. *Let* X_n*,* $n \geq 0$ *and* X *be random variables.*

(i) *We say that* X_n **converges in probability** *to* X*,* $X_n \xrightarrow{P} X$*, if* $\forall\, \varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

(ii) *We say that* X_n **converges almost surely** *to* X*,* $X_n \xrightarrow{a.s.} X$*, if* $\forall\, \varepsilon$

$$\lim_{N\to\infty} P(|X_n - X| < \varepsilon\ \forall n \geq N) = 1.$$

It it also useful to note that almost sure convergence $X_n \xrightarrow{a.s.} X$ implies convergence in probability $X_n \xrightarrow{P} X$.

Before stating the Martingale Convergence Theorem, the following inequalities are also pertinent in its proof.

Proposition 10 (Schwartz's Inequality). *Let X, Y be random variables s.t. $E\big(|X|^2\big) < \infty$ and $E\big(|Y|^2\big) < \infty$. Then*

$$E\big(|XY|\big) \leq E\big(|X|^2\big)^{1/2} \cdot E\big(|Y|^2\big)^{1/2}.$$

Proposition 11 (Doob-Kolmogorov Martingale Inequality). *If $\{X_n\}_{n\geq 0}$ is a martingale and $E\big(X_n^2\big) < \infty$ for all n, then*

$$P(\max\{X_0, X_1, \ldots, X_n\} \geq \epsilon) \leq \frac{E\big(X_n^2\big)}{\epsilon^2}.$$

Theorem 1 (Martingale Convergence Theorem). *If $\{X_n\}_{n\geq 0}$ is a martingale and $\exists\, M < \infty$ s.t. $E(|X_n|) \leq M$ for all n, then there exists a random variable X s.t.*

$$X_n \xrightarrow{a.s.} X.$$

2.5 Law of Large Numbers and Central Limit Theorem for Martingales

For the strong law of large numbers for martingales, let $\{X_n\}_{n\geq 0}$ be a sequence of random variables such that, with probability 1,

$$E(X_0) = 0, \quad E(X_{n+1}|X_0, X_1, \ldots, X_n) = 0.$$

Let $S_n = X_0 + \ldots + X_n$. The sequence $\{S_n\}_{n\geq 0}$ forms a martingale:

$$E(S_{n+1}|S_0, \ldots, S_n) = S_n.$$

Theorem 2 (Strong Law of Large Numbers for Martingales). *Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables s.t. $E(X_0) = 0$ and $E(X_{n+1}|X_0, X_1, \ldots, X_n) = 0$. Let $S_n = X_0 + \ldots + X_n$. If $\sum_{n=1}^{\infty} \frac{X_n^2}{n^2} < \infty$, then*

$$\frac{S_n}{n} \xrightarrow{a.s.} 0.$$

Theorem 3 (Central Limit Theorem for Martingales). *Let $\{X_n\}_{n\geq 0}$ be a martingale and suppose*

$$E(X_{n+1} - X_n|X_0, \ldots, X_n) = 0, \quad |X_{n+1} - X_n| \leq k$$

with probability 1 for some fixed k and for all n. Suppose $|X_0| \leq k$ with probability 1. Define $\sigma_n^2 = E\big((X_{n+1} - X_n)^2 | X_0, \ldots, X_n\big)$ *and let* $\tau_\nu = \min\left\{n : \sum_{i=0}^{n} \sigma_i^2 \geq \nu\right\}$. *Then as* $\nu \to \infty$

$$\frac{X_{\tau_\nu}}{\sqrt{\nu}} \xrightarrow{\mathcal{D}} X \sim N(0,1).$$

These two theorems will be revisited in a connection to controlled Markov chains.

3 An Overview of Adaptive Control of Markov Chains

3.1 Introduction to Controlled Markov Chains

Markov chains are desirable stochastic processes because it is not necessary to store past information to make predictions. However, the trajectory of the chain might not naturally lead to the desired state. This is where control theory comes into play: introducing control actions to the process ideally allows us to steer the chain in our desired direction.

In this paper, we are interested in a finite controlled Markov chain. The model is as follows: Let the sequence of random variables $\{X_n\}_{n\geq 0}$ be the state variables from a finite state space I. Let the sequence of random variables $\{u_n\}_{n\geq 0}$ be the control actions, defined as functions of the state variables s.t.

$$U(X_n) = u_n.$$

Then the transition probabilities of this controlled Markov chain are of the form

$$P(X_{n+1} = j | X_n = i) = p(i, j; u_n), \quad j \in I$$

where at time n, X_n is observed as state $i \in I$, and based on this information u_n is selected as the control action from a prespecified set U.

In [4], Kumar and Varaiya have presented a simple example of a controlled Markov chain. I will present a similar example, which also demonstrates the usefulness of controlled Markov chains.

Example 9. Consider a system whose condition at time n, which is described as the state X_n, can take the values 1 or 2 such that $X_n = 1$ or $X_n = 2$ depending on whether the system is in an operational condition or a failed condition. Without control actions, the behavior of this system

is autonomous. Suppose the system is operational at time n ($X_n = 1$) and it has probability p of staying operational at the next time ($X_{n+1} = 1$) and probability $1 - p$ of failing at the next time ($X_{n+1} = 2$). Suppose p only depends on the current state at time n. Lastly, let us say once the system has failed, it remains failed, so if $X_n = 2$, then $X_{n+1} = 2$ with probability 1.

Note, $\{X_n\}_{n\geq 0}$ is a Markov chain with transition probability matrix

$$\mathcal{P} = \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix}.$$

Let us introduce two control actions, u_n^1 and u_n^2. Let u_n^1 denote the intensity of usage of the system at time n, taking values 0 for not used, 1 for lightly used, and 2 for heavily used. We will say that the higher intensity of usage, the more likely the system is to fail. Let u_n^2 denote the intensity of maintenance performed on the system, taking values 0 for low amounts of maintenance and 1 for high amounts of maintenance. The more maintenance performed on the system, the less likely the system is to fail. Let $u_n = (u_n^1, u_n^2)$. Now we have the controlled transition probabilities

$$\begin{aligned}
&P(X_{n+1} = 1 | X_n = 1, X_{n-1}, \ldots; u_n, u_{n-1}, \ldots) = p_1(u_n^1) - p_2(u_n^2) \\
&P(X_{n+1} = 2 | X_n = 1, X_{n-1}, \ldots; u_n, u_{n-1}, \ldots) = 1 - [p_1(u_n^1) - p_2(u_n^2)] \\
&P(X_{n+1} = 1 | X_n = 2, X_{n-1}, \ldots; u_n, u_{n-1}, \ldots) = p_2(u_n^2) \\
&P(X_{n+1} = 2 | X_n = 2, X_{n-1}, \ldots; u_n, u_{n-1}, \ldots) = 1 - p_2(u_n^2).
\end{aligned}$$

In matrix form,

$$\mathcal{P}(u) = \begin{bmatrix} p_1(u_n^1) - p_2(u_n^2) & 1 - [p_1(u_n^1) - p_2(u_n^2)] \\ p_2(u_n^2) & 1 - p_2(u_n^2) \end{bmatrix}.$$

Now, with controlled Markov chains, the current state is observed, and we choose a control action based upon this observation. We do so by a control feedback law taking in the observation and putting out the best control choice, say $\phi(X_n) = u_n$. Let us say $\phi(1) = (2, 0)$ and $\phi(2) = (0, 1)$.

Then we have the transition probability matrix

$$\mathcal{P}^{\phi} = \begin{bmatrix} p_1(2) - p_2(0) & 1 - [p_1(2) - p_2(0)] \\ p_2(1) & 1 - p_2(1) \end{bmatrix}.$$

From this, we can see that changing the feedback law, i.e. changing the amount of usage of the system and maintenance on the system, we can affect the probability of the system remaining operational or failing.

3.2 Controlled Markov Chains and Martingales

In [5], Mandl describes some results of the reward of the controlled Markov chain from theory on martingales. The reward of the previously described chain up to time N is given by

$$C_N = \sum_{n=0} c(i, j; u_n).$$

Mandl then sets out to construct a martingale as a basis on which the Law of Large Numbers and Central Limit Theorem of controlled Markov chains can be derived.

To construct the martingale, Mandl defines Θ a real number, w_i, w_j auxiliary constants, and

$$\varphi(i, u_n) = \sum_j p(i, j; u_n)\big[c(i, j; u_n) + w_j\big] - w_i - \Theta.$$

Then Mandl sets

$$Y_n = c(i, j; u_n) - \Theta + w_j - w_i - \varphi(i, u_n) \implies E(Y_n | X_0, \ldots, X_n) = 0,$$

and

$$B_N = \sum_{n=0}^{N-1} Y_n$$

is a martingale with respect to $\{X_0, \ldots, X_N\}$.

The Strong Law of Large Numbers for martingales (section 2.5, theorem 2) yields

$$\lim_{N \to \infty} \frac{B_N}{N} = 0 \quad \text{a.s.}$$

To obtain the Law of Large Numbers for controlled Markov chains, one last property is required.

Property 2. *The states $i \in I$ that are recurrent for a Markov chain with transition matrix $[p(i,j;U(i))]$, $i,j \in I$ form only one irreducible set.*

Then we can find constants Θ, w_i s.t. $\varphi(i,U(i)) = 0$ for $i \in I$.

Theorem 4 (Law of Large Numbers for Controlled Markov Chains)**.** *Suppose $\varphi(i,U(i)) = 0$ for $i \in I$. If*

$$\lim_{n\to\infty} u_n - U(X_n) = 0 \quad a.s.,$$

then

$$\lim_{N\to\infty} \frac{C_N}{N} = \Theta \quad a.s.$$

From this theorem, we also have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} c_2(X_n, u_n) = \sigma^2,$$

where

$$c_2(i,u_n) = \sum_j p(i,j;u_n)\big[c(i,j;u_n) - \Theta + w_j - w_i\big]^2 \quad i,j \in I$$

and σ^2 is attainable with auxiliary constants w_{2i}, $i \in I$.

Then with regard to the martingale B_N, the following theorem is obtained.

Theorem 5 (Central Limit Theorem for Controlled Markov Chains)**.** *Suppose $\sigma^2 > 0$, $\varphi(i,U(i)) = 0$ for $i \in I$, and*

$$\sum_j p(i,j;U(i))\big[(c(i,j;U(i)) - \Theta)^2 + 2\big(c(i,j;U(i)) - \Theta\big)w_j + w'_{2j}\big] - w'_{2i} - \sigma^2 = 0,$$

where $w'_{2i} = w_{2i} + w_i^2$ for $i \in I$. If

$$\lim_{n\to\infty} u_n - U(X_n) = 0 \quad a.s.,$$

then for $N \to \infty$

$$\frac{C_N - N\Theta}{\sigma\sqrt{N}} \sim N(0,1).$$

3.3 Adaptive Control of Markov Chains

In a perfect world, all parameters of a Markov chain would be known in order to choose the best control at each step. However, in many cases the behavior of a Markov chain depends on some unknown parameters, so the controls are adjusted based on estimates of the unknown parameters from observations of the process at time n.

In the paper [6], Mandl considered a controlled Markov chain $\{X_n\}_{n\geq 0}$, taking values in a finite set I with transition probabilities that depend upon the control actions u_n at time n and parameter α:

$$P(X_{n+1} = j | X_n = i) = p(i, j; u_n, \alpha).$$

At each time n, X_n is observed, and u_n is selected based on X_n. The parameter α has the constant true value α°, which is unknown.

Mandl constructs the adaptive control law based upon maximum likelihood estimation of the unknown parameter. The maximum likelihood estimate at time n is denoted α_n.

For the main result, Mandl establishes the following assumptions:

(A1) For $i, j \in I$ either

$$p(i, j; u, \alpha) > 0 \ \ \forall u, \alpha \quad \text{or} \quad p(i, j; u, \alpha) = 0 \ \ \forall u, \alpha.$$

(A2) **Identifiability Condition**: For each $\alpha \neq \alpha'$, there exists $i \in I$ so that

$$[p(i, 1; u, \alpha), \ldots, p(i, I; u, \alpha)] \neq [p(i, 1; u, \alpha'), \ldots, p(i, I; u, \alpha')],$$

for all $u \in U$.

The identifiability condition is very strict, but allows Mandl to achieve the following desired result:

Theorem 6 (Mandl)**.** *Let* $(A1), (A2)$ *hold. Then as* $n \to \infty$*,*

$$\alpha_n \xrightarrow{a.s.} \alpha^\circ.$$

In other words, the Identifiability Condition yielded the result that the sequence of maximum likelihood estimates α_n would converge almost surely to the true parameter α°. This is a best-case scenario, however the Identifiability Condition has been argued to be impractical. Counterexamples to this condition are given in [1] and [3].

3.3.1 Finite Parameter Set

In [1], Borkar and Varaiya wish to consider the same controlled Markov chain as Mandl, but without the strict Identifiability Condition. They similarly consider the controlled Markov chain $\{X_n\}_{n\geq 0}$, taking values in a finite set I with transition probabilities that depend upon the control actions u_n at time n and parameter α:

$$P(X_{n+1} = j | X_n = i) = p(i, j; u_n, \alpha).$$

The parameter α has the constant true value α°, which is unknown.

Borkar and Varaiya also construct the adaptive control law from maximum likelihood estimation of the unknown parameter. The maximum likelihood estimate α_n is then used in selecting the control action $u_n = \phi(\alpha_n, X_n)$, where $\phi(\alpha, \cdot)$ is a stationary control law, and the corresponding likelihood functions $L_n(\alpha)$ at time n are noted to be positive martingales. Their main objective is to analyze the asymptotic behavior of α_n and u_n, and specifically when the identifiability condition proposed by Mandl may not hold.

For their main results, Borkar and Varaiya assume that α°, the true parameter, is known to belong to a finite parameter set A. There are two further assumptions:

(A3) There exists $\varepsilon > 0$ s.t. for every i, j, either

$$p(i, j; u, \alpha) > \varepsilon \ \ \forall u, \alpha \quad \text{or} \quad p(i, j; u, \alpha) = 0 \ \ \forall u, \alpha.$$

(A4) For every i, j there is a sequence $i_1, \ldots, i_r$ s.t. $\forall u, \alpha$,

$$p(i_{s-1}, j_s; u, \alpha) > 0, \quad s = 1, \ldots, r,$$

where $i_0 = i,\ i_{r+1} = j$.

Note, (A3) is adapted from (A1) in [6]. (A4) guarantees that the Markov chain generated from transition probabilities $p(i, j; \phi(\alpha, i), i)$ has a single recurrent class, which is a necessary assumption for identification.

The main result Borkar and Varaiya achieved in this paper is stated in the following theorem:

Theorem 7 (Borkar-Varaiya)**.** *There is a set W of zero measure, a random variable $\alpha^* \in A$, and a finite random time N s.t. for $\omega \notin W$, $n \geq N(\omega)$,*

$$\alpha_n(\omega) = \alpha^*(\omega), \qquad u_n(\omega) = \phi(\alpha^*(\omega), X_n(\omega)),$$
$$p((i, j; \phi(\alpha^*(\omega), i), \alpha^*(\omega)) = p(i, j; \phi(\alpha^*(\omega), i), \alpha^\circ),$$

for all $i, j \in I$.

Written more generally, they determined α_n converges almost surely to a random variable α^* s.t.

$$p(i, j; \phi(\alpha^*, i), \alpha^*) = p(i, j; \phi(\alpha^*, i), \alpha^\circ), \quad \forall\, i, j.$$

Thus, asymptotically, the transition probabilities of their adaptive control system are the same whether the parameter is α^* or the true α°, meaning the two parameters are indistinguishable.

3.3.2 Compact Parameter Set

Now in [3], Kumar's main objective is to make Borkar and Varaiya's assumption of a finite parameter set more general, while still considering when Mandl's identifiability condition may not hold. To do this, Kumar considers a compact parameter set.

Kumar considers the same controlled Markov chain $\{X_n\}_{n\geq 0}$ with transition probabilities $p(i, j; u_n, \alpha)$ that depend upon the control actions u_n and parameter α. The parameter α has the unknown true value α°. Kumar assumes that α° belongs to a compact parameter set A instead of the finite set assumed by Borkar and Varaiya.

Similarly to Borkar and Varaiya, at each time n Kumar takes maximum likelihood estimate α_n of parameter α°, then applies the control action $u_n = \phi(\alpha_n, X_n)$.

For the main results, Kumar establishes the following assumptions:

(A5) The transition probabilities $p(\cdot,\cdot;\cdot,\cdot)$ and the control law $\phi(\cdot,\cdot)$ are continuous.

(A6) For every i, j, either

$$p(i,j;u,\alpha) > 0 \ \ \forall u,\alpha, \quad \text{or} \quad p(i,j;u,\alpha) = 0 \ \ \forall u,\alpha.$$

(A7) For every i, j, there exists a sequence $i = i_0, i_1, \ldots, i_r = j$ s.t.

$$p(i_{s-1}, i_s; u_s, \alpha) > 0 \quad \forall s = 1, \ldots, r.$$

Note, (A6) mimics (A1) from [6] and (A3) from [1], and (A7) mimics (A4) from [1].

The main results Kumar achieved in this paper is stated in the following theorem:

Theorem 8 (Kumar)**.** *There exists a null set W, $P(W) = 0$ s.t. if for some $\omega \in W^c$ the control converges to ψ, then*

$$p(i,j;\psi(i),\alpha^*) = p(i,j;\psi(i),\alpha^\circ)$$

for all i, j and every limit point α^* of $\{\alpha_n\}_{n\geq 0}$.

As an important consequence, if $\lim_{n\to\infty} \alpha_n(\omega) = \alpha^*$, then

$$p(i,j;\phi(\alpha^*,i),\alpha^*) = p(i,j;\phi(\alpha^*,i),\alpha^\circ) \quad \forall i,j.$$

Thus, if the parameter estimates do converge, or more generally if only the control laws converge to some feedback law ψ, then under ψ the transition probabilities corresponding to any limit point of $\{\alpha_n\}$ are the true transitions probabilities (corresponding to α°).

4 Parameter Estimation in a Linear Case

In this paper, we will extend the results in [8] from a linear case to a quadratic case. Before extending to the quadratic case, however, this section will survey Sagalovsky's results and methods in the linear case.

In [8], the main objective is to build upon Borkar and Varaiya's consideration when Mandl's Identifiability Condition may not hold. To start, Sagalovsky establishes the same setup as Borkar

and Varaiya in [1]. So, Sagalovsky considers the controlled Markov chain $\{X_n\}_{n\geq 0}$ taking values in a finite set I with transition probabilities that depend upon the control actions u_n selected at time n based on previous observations and parameter α:

$$P(X_{n+1} = j | X_n = i; u_n, \alpha) = p(i, j; u_n, \alpha).$$

Again, the parameter α has the true value α°, which is unknown.

Then, Sagalovsky considers a model where the transition probabilities depend linearly on α:

$$p(i, j; u_n, \alpha) = a(i, j; u_n)\alpha + b(i, j; u_n),$$

where a, b are known real functions. Assuming α° belongs to a bounded interval $A = (\underline{\alpha}, \overline{\alpha})$, Sagalovsky also estimates the unknown parameter by its maximum likelihood estimate α_n. To do so, the likelihood of a given α at time n is defined as

$$P(X_0, \ldots, X_n | X_0, \alpha) = \prod_{m=0}^{n-1} p(X_m, X_{m+1}; u_m, \alpha)$$

and the log-likelihood as

$$L_n(\alpha) = \sum_{m=0}^{n-1} \log p(X_m, X_{m+1}; u_m, \alpha) = \sum_{m=0}^{n-1} \log \big[a(X_m, X_{m+1}; u_m)\alpha + b(X_m, X_{m+1}; u_m)\big].$$

Let $a_m = a(X_m, X_{m+1}; u_m)$ and $b_m = b(X_m, X_{m+1}; u_m)$, so we have the simpler log-likelihood function

$$L_n(\alpha) = \sum_{m=0}^{n-1} \log \big[a_m\alpha + b_m\big].$$

Let $a_m^j = a(X_m, j; u_m)$ and $b_m^j = b(X_m, j; u_m)$. Then, Sagalovsky defines the maximum likelihood estimate (MLE) α_n at time n to be the element of $\overline{A}$ (the closure of A) s.t. $L_n(\alpha_n) \geq L_n(\alpha)$ $\forall \alpha \in \overline{A}$. The MLE is found over the closure of A to ensure such an α_n exists.

Note

$$L_n'(\alpha) = \sum_{m=0}^{n-1} \frac{a_m}{a_m\alpha + b_m}$$
$$L_n''(\alpha) = -\sum_{m=0}^{n-1} \left(\frac{a_m}{a_m\alpha + b_m} \right)^2 \leq 0.$$

Since the second derivative is nonpositive, there exists at least one $\hat{\alpha}_n$ maximizing $L_n(\alpha)$ for $\alpha \in \overline{A}$. This $\hat{\alpha}_n$ is unique if $\exists m < n$ s.t. $a_m \neq 0$. In what follows Sagalovsky assumes this m exists, so $\hat{\alpha}_n$ is called α_n as it is unique.

To find the maximizing α_n, Sagalovsky considers the sign of $L_n'(\alpha)$.

For the main results, Sagalovsky establishes the following assumptions:

(A8) $\forall i, j$, either

$$p(i, j; u, \alpha) \geq \epsilon > 0 \ \ \forall u, \alpha \quad \text{or} \quad p(i, j; u, \alpha) = 0 \ \ \forall u, \alpha.$$

(A9) $u_n = \phi(\alpha_n, x_n)$ and $a(i, j; \phi(\alpha, i))$ are continuous in α for every i, j.

(A10) For every i, j, there exists a sequence $i = i_0, i_1, \ldots, i_r, i_{r+1} = j$ s.t.

$$p(i_{s-1}, i_s; u_s, \alpha) > 0 \quad \forall s = 1, \ldots, r+1.$$

Notice again the similarities between (A8) and (A1) [6], (A3) [1], (A6) [3], and between (A10) and (A4) [1], (A7) [3].

Sagalovsky achieved the following important result:

Theorem 9 (Sagalovsky)**.** *Under (A8), except for a P-null set of realizations, if the sequence* $\{\alpha_n\}$ *of MLE's has an accumulation point* $\alpha^* \neq 0$*, then*

$$\sum_{m=0}^{\infty} \left\{ \sum_{j \in I} a_m^{j^2} \right\} < \infty.$$

This implies $a_m^j \to 0$ for each $j = 1, \ldots, I$. Sagalovsky notes that this gives the intuitive feeling that, for α_n not to converge to the true value, the transitions should give increasingly less

information on α as n grows as the term $a(i,j;u)$ is the term that relates to α.

Theorem 10 (Sagalovsky)**.** *Under (A8), for all realizations not in a P-null set,*

$$\alpha_n \to \alpha^* \quad \text{as } n \to \infty.$$

So under (A8), $\{\alpha_n\}$ converges almost surely, though not necessarily to the true parameter α°. Lastly, under (A8)-(A10), Sagalovsky states that if $\{\alpha_n\}$ has an accumulation point α^*, then

$$p(i,j;\phi(\alpha^*,i),\alpha^*) = p(i,j;\phi(\alpha^*,i),\alpha^\circ) \quad \forall i,j.$$

Thus, asymptotically, the transition probabilities of this adaptive control system are the same whether the parameter is α^* or the true α°, meaning the two parameters are indistinguishable (similar to Borkar and Varaiya).

Sagalovsky also explains that the above result with the case that $\{\alpha_n\}$ does not converge to the true parameter actually yields a rule to modify the control law ϕ as to guarantee convergence to the true parameter α°: if necessary, modify $\phi(\alpha_n, X_n)$ so that

$$\sum_{m=0}^{\infty} \left\{ \sum_{j \in I} a_m^{j^2} \right\} = \infty.$$

5 Extension to a Quadratic Case

In [7], Pasik-Duncan extends Sagalovsky's results in the linear case to a two-dimensional linear case, and leaves the extension (using similar methods) to the quadratic case to the reader. In this paper, the objective is to present these methods to extend the results in the last section to a quadratic case.

5.1 Model

We consider a Markov chain $\{X_n\}_{n\geq 0}$ which takes values from a finite state space $I = \{1, 2, \ldots, I\}$. The transition probabilities $P(X_{n+1} = j | X_n = i)$ are assumed to depend on an unknown real parameter α. These transition probabilities are also affected by the control action u_n selected at

time n based on the previous observations on the direction of the chain. Denote

$$P(X_{n+1} = j | X_n = i; u_n, \alpha) = p(i, j; u_n, \alpha).$$

Suppose parameter α has the unknown true constant value α°. The specific model we will consider in this section is quadratic, where the transition probabilities depend quadratically on α, i.e.

$$p(i, j; u_n, \alpha) = a(i, j; u_n)\alpha^2 + b(i, j; u_n)\alpha + c(i, j; u_n), \tag{1}$$

where $a(\cdot, \cdot; \cdot), b(\cdot, \cdot; \cdot), c(\cdot, \cdot; \cdot)$ are known real functions.

The control u_n only depends on the procession of the chain from time 0 up to and including the current state. It is assumed this procession is observed exactly, so defining the σ-algebra generated by $X_0, X_1, \ldots, X_m$

$$\mathcal{F}_m = \sigma\{X_0, X_1, \ldots, X_m\}$$

allows u_n to be $\mathcal{F}_m$-measurable for $m = 0, 1, \ldots$. We will write $f \in \mathcal{F}_m$ to denote $\mathcal{F}_m$-measurability of a given function f.

Recall, the true value α° for parameter α is unknown. However, we assume it is known that α° belongs to an interval J with endpoints $\underline{\alpha} < \overline{\alpha}$. For simplicity, assume $\alpha^\circ = 0$, and $\underline{\alpha} < -1$ and $\overline{\alpha} > 1$, that way $-1, 1 \in J$.

We want to control the procession of the chain with our choices of controls u_n. To choose the best control u_n at time n, we need to estimate α°. As in the preceding papers discussed in the previous sections, we will estimate α° using maximum likelihood estimation. We define the likelihood of a given α at time n as

$$P(X_0, X_1, \ldots, X_n | X_0; u_0, u_1, \ldots, u_{n-1}, \alpha) = \prod_{m=0}^{n-1} p(X_m, X_{m+1}; u_m, \alpha)$$

and the log-likelihood

$$\begin{aligned} L_n(\alpha) &= \sum_{m=0}^{n-1} \log p(X_m, X_{m+1}; u_m, \alpha) \\ &= \sum_{m=0}^{n-1} \log \left[a(X_m, X_{m+1}; u_m)\alpha^2 + b(X_m, X_{m+1}; u_m)\alpha + c(X_m, X_{m+1}; u_m) \right], \end{aligned}$$

$\alpha \in J$. Let $a_m = a(X_m, X_{m+1}; u_m)$, $b_m = b(X_m, X_{m+1}; u_m)$, and $c_m = c(X_m, X_{m+1}; u_m)$. Further, let $a_m^j = a(X_m, j; u_m)$, $b_m^j = b(X_m, j; u_m)$, and $c_m^j = c(X_m, j; u_m)$. Note $a_m, b_m, c_m \in \mathcal{F}_{m+1}$ and $a_m^j, b_m^j, c_m^j \in \mathcal{F}_m$ for $j = 1, \ldots, I$. Now we have the simpler log-likelihood function

$$L_n(\alpha) = \sum_{m=0}^{n-1} \log \left[a_m\alpha^2 + b_m\alpha + c_m \right],$$

$\alpha \in J$. Now, we will define the maximum likelihood estimate (MLE) of α° at time n as α_n, the element in $\overline{J}$ s.t.

$$L_n(\alpha_n) \geq L_n(\alpha) \qquad\qquad \forall\, \alpha \in \overline{J}.$$

Since we defined the MLE over the closure of J, $\overline{J}$, we can guarantee the existence of at least one such MLE. If there is more than one element that satisfies our maximum likelihood criteria, we can choose α_n to be the smallest such element.

5.2 Assumptions and Formulation

In this section, we will outline the assumptions and their contributions used to achieve our convergence results. Recall for the transition probabilities of a Markov chain we have

$$\sum_{j=1}^{I} p(i, j; u, \alpha) = 1 \qquad\qquad \forall\, \alpha, u.$$

Then from our defined model (1), it follows that

$$\sum_{j=1}^{I} a(i, j; u) = 0, \quad \sum_{j=1}^{I} b(i, j; u) = 0, \quad \sum_{j=1}^{I} c(i, j; u) = 1 \qquad\qquad \forall\, i, u. \tag{2}$$

Further, since $0 \leq p(i,j;u,\alpha) \leq 1 \ \forall\, i,j,u,\alpha$, and α can take on the values $-1, 1$, we have that

$$0 \leq |a(i,j;u)| \leq c(i,j;u) \leq 1, \quad 0 \leq |b(i,j;u)| \leq c(i,j;u) \leq 1, \tag{3}$$

$$0 \leq |a(i,j;u) - b(i,j,u)| \leq c(i,j;u) \leq 1 \tag{4}$$

$\forall\, i,j \in I$ and $\forall\, u$.

We will now make the following assumption also employed in [6], [1], [3], and [8].

(A11) For all $i,j \in I$, either

$$p(i,j;u,\alpha) = 0 \ \ \forall\, u,\alpha \quad \text{or} \quad p(i,j;u,\alpha) \geq K > 0 \ \ \forall u,\alpha.$$

This assumption allows us to determine topology of the chain, which should be known and should not be altered by the choice of control u. The only other true restriction is the uniform lower bound provided by K.

We will not consider the P-null set of realizations where transitions with 0 probability occur. Hence, under (A11), $L_n(\alpha)$ is infinitely differentiable in α and we can compute

$$L_n'(\alpha) = \sum_{m=0}^{n-1} \frac{2a_m\alpha + b_m}{a_m\alpha^2 + b_m\alpha + c_m}, \tag{5}$$

$$L_n''(\alpha) = -\sum_{m=0}^{n-1} \frac{2a_m^2\alpha^2 + 2a_m b_m \alpha + b_m^2 - 2a_m c_m}{(a_m\alpha^2 + b_m\alpha + c_m)^2}. \tag{6}$$

To ensure that the likelihood function $L_n(\alpha)$ achieves its maximum and its second derivative is a nonpositive function, let us make the following assumptions:

(A12) For all m, either

$$a_m, b_m \geq 0 \quad \text{or} \quad a_m, b_m \leq 0.$$

(A13) For all m,

$$a_m \cdot c_m < 0.$$

From (6), it follows that there exists at least one $\hat{\alpha}_n$ maximizing $L_n(\alpha)$ for $\alpha \in \overline{J}$ and that this $\hat{\alpha}_n$ is unique if $\exists m < n$ s.t. $a_m \neq 0$, $b_m \neq 0$, and $c_m \neq 0$. It could be the case for some realization

$a_m = b_m = c_m = 0$ for all m. Then $L_m(\alpha)$ would be constant in α and we would take $\alpha_m = \underline{\alpha}$ for all m. In this case the results we claim in this section will hold. For this reason we consider in what follows that $\exists m$ s.t. $a_m \neq 0$, $b_m \neq 0$, and $c_m \neq 0$ and that n is larger than this m.

To find the maximizing $\hat{\alpha}_n$, we will consider the sign of $L_n'(\alpha)$. Define for $m = 0, 1, \ldots$

$$D_m(\alpha) = \frac{2a_m\alpha + b_m}{a_m\alpha^2 + b_m\alpha + c_m} \tag{7}$$

and note

$$L_n'(\alpha) = \sum_{m=0}^{n-1} D_m(\alpha), \quad \text{and} \quad D_m(\alpha) \in \mathcal{F}_{m+1}. \tag{8}$$

Define also

$$E_m(\alpha) = E(D_m(\alpha)|\mathcal{F}_m) = \sum_{j=1}^{I} \left(\frac{2a_m^j\alpha + b_m^j}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j} \right) \cdot c_m^j, \tag{9}$$

$$V_m(\alpha) = \mathrm{Var}(D_m(\alpha), \mathcal{F}_m), \tag{10}$$

where $c_m^j = p(X_m, j; u, \alpha^0 = 0)$. Now we have

$$E_m(\alpha^0 = 0) = \sum_{j=1}^{I} \left(\frac{b_m^j}{c_m^j} \right) \cdot c_m^j = \sum_{j=1}^{I} b_m^j = 0 \tag{11}$$

by (2), and

$$\frac{dE_m(\alpha)}{d\alpha} = -\sum_{m=0}^{n-1} \left(\frac{2(a_m^j)^2\alpha^2 + 2a_m^j b_m^j \alpha + (b_m^j)^2 - 2a_m^j c_m^j}{(a_m^j\alpha^2 + b_m^j\alpha + c_m^j)^2} \right) \cdot c_m^j \leq 0, \tag{12}$$

in which case we have

$$\alpha < 0 \implies E_m(\alpha) \geq 0, \qquad \alpha > 0 \implies E_m(\alpha) \leq 0. \tag{13}$$

We will consider the behavior of $L_n'(\alpha)$ at a given $\alpha < 0$. A symmetric argument can be used for $\alpha > 0$. Fixing α, for simplicity we will write D_m, E_m for $D_m(\alpha)$, $E_m(\alpha)$, etc.

Define

$$Y_m = D_m - E_m \in \mathcal{F}_{m+1}, \tag{14}$$

with $E(Y_m|\mathcal{F}_m) = 0$, and let

$$A_n = \sum_{m=0}^{n-1} E_m \in \mathcal{F}_{n-1}, \tag{15}$$

$$V_m = \mathrm{Var}(Y_m|\mathcal{F}_m), \tag{16}$$

$$M_n = \sum_{m=0}^{n-1} Y_m \in \mathcal{F}_n. \tag{17}$$

Note, M_n is a square integrable martingale. Also, let

$$\overline{M}_n = \sum_{m=0}^{n-1} V_m. \tag{18}$$

Then we have $M_n^2 - \overline{M}_n$ is a martingale. Note

$$L'_n = \sum_{m=0}^{n-1} (Y_m + E_m) = M_n + A_n \tag{19}$$

Now, some computations are needed in order to proceed further:

$$E_m = \sum_{j=1}^{I} \frac{\left(2a_m^j\alpha + b_m^j\right)\cdot c_m^j}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j} \tag{20}$$

$$= \sum_{j=1}^{I} \left(2a_m^j\alpha + b_m^j\right)\cdot \frac{c_m^j}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j} \tag{21}$$

$$= \sum_{j=1}^{I} \left(2a_m^j\alpha + b_m^j\right)\left(1 - \frac{(a_m^j\alpha^2 + b_m^j\alpha)}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j}\right) \tag{22}$$

$$= 2\sum_{j=1}^{I} a_m^j\alpha + \sum_{j=1}^{I} b_m^j - \sum_{j=1}^{I} \frac{2(a_m^j)^2\alpha^3 + 3a_m^j b_m^j\alpha^2 + (b_m^j)^2\alpha}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j} \tag{23}$$

$$= -\sum_{j=1}^{I} \frac{2(a_m^j)^2\alpha^3 + 3a_m^j b_m^j\alpha^2 + (b_m^j)^2\alpha}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j} \tag{24}$$

$$= -\alpha\sum_{j=1}^{I} \frac{2(a_m^j)^2\alpha^2 + 3a_m^j b_m^j\alpha + (b_m^j)^2}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j}, \tag{25}$$

where we used (2). From our assumptions on a_m, b_m, c_m and that $\alpha < 0$, we have that $E_m \geq 0$.

Let us estimate the ratio V_m/E_m:

$$\frac{V_m}{E_m} \leq \frac{\displaystyle\sum_{j=1}^{I} \frac{\left(2a_m^j\alpha + b_m^j\right)^2 c_m^j}{(a_m^j\alpha^2 + b_m^j\alpha + c_m^j)^2}}{\displaystyle\sum_{j=1}^{I} \frac{\left(2a_m^j\alpha + b_m^j\right)^2 c_m^j}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j}} \leq \frac{\displaystyle\frac{1}{K}\sum_{j=1}^{I} \frac{(2a_m^j\alpha)^2}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j}}{\displaystyle\sum_{j=1}^{I} \frac{\left(2a_m^j\alpha + b_m^j\right)^2 |\alpha|}{a_m^j\alpha^2 + b_m^j\alpha + c_m^j}} = \frac{1}{K|\alpha|}. \tag{26}$$

So, we have

$$V_m \leq \frac{E_m}{K|\alpha|} \tag{27}$$

The increasing process $\overline{M_n}$ plays an important role in defining the behavior of the martingale M_n. We have the following lemma from [7]:

Lemma 1. *For all realizations not in a P-null set,*

$$\lim_{n\to\infty} \overline{M}_n < \infty \implies M_n \to M < \infty \quad \text{as } n \to \infty \tag{28}$$

$$\lim_{n\to\infty} \overline{M}_n < \infty \implies \frac{M_n}{\overline{M}_n} \to 0 \quad \text{as } n \to \infty. \tag{29}$$

Proof.

$$\lim_{n\to\infty} \overline{M}_n < \infty \implies \frac{L_n'(\alpha)}{A_n} = 1 + \frac{M_n}{A_n} \to 1 + \frac{M}{\lim_{n\to\infty} A_n} \to 1 \quad \text{as } n \to \infty,$$
$$\lim_{n\to\infty} \overline{M}_n = \infty \implies \frac{L_n'(\alpha)}{A_n} = 1 + \frac{M_n}{A_n} \to 1 \quad \text{as } n \to \infty,$$

using (27).

□

Theorem 11. *For all realizations not in a P-null set*

$$A_n = \sum_{m=0}^{n-1} E_m \to \infty \implies \frac{L_n'(\alpha)}{A_n} \to 1 \quad \text{as } n \to \infty$$

for $\alpha < 0$.

From (A11) and $a_m^j\alpha^2 + b_m^j\alpha + c_m^j \leq 1$ for each $j \in I$, it follows that for $\alpha < 0$,

$$\sum_{j=1}^{I} \left[2(-\alpha)^3(a_m^j)^2 - 3\alpha^2 a_m^j b_m^j + (-\alpha)(b_m^j)^2\right] \leq E_m(\alpha) \tag{30}$$

$$\leq \frac{\sum_{j=1}^{I} \left[2(-\alpha)^3(a_m^j)^2 - 3\alpha^2 a_m^j b_m^j + (-\alpha)(b_m^j)^2\right]}{K}. \tag{31}$$

Then, note that for all $\alpha < 0$,

$$\sum_{m-0}^{n-1}\sum_{j=1}^{I} \left[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2\right] \to \infty \implies L_n'(\alpha) \to \infty \quad \text{as } n \to \infty. \tag{32}$$

An analogous argument can be made for $\alpha > 0$. This leads us to the following corollary.

Corollary 1. *For every realization except on a P-null set*

$$\sum_{m-0}^{n-1}\sum_{j=1}^{I}\left[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2\right] \to \infty \quad \text{as } n\to\infty \implies \begin{cases} L_n'(\alpha)\to\infty & \text{as } n\to\infty \\ L_n'(\alpha)\to-\infty & \text{as } n\to\infty \end{cases} \tag{33}$$

This leads us to the following theorem:

Theorem 12. *If the sequence $\{\hat{\alpha}_n\}$ of maximum likelihood estimates has an accumulation point $\alpha^* \neq 0$, then for every realization except on a P-null set*

$$\sum_{m=0}^{\infty}\left(\sum_{j=1}^{I}\left[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2\right]\right) < \infty \tag{34}$$

Proof. Consider a realization where (33) holds and the sequence $\{\hat{\alpha}_n\}$ has an accumulation point $\alpha^* < 0$ (symmetric argument for $\alpha^* > 0$). Then there exists a subsequence $\{n_k\}$ s.t.

$$\lim_{k\to\infty}\hat{\alpha}_{n_k} = \alpha^* < 0.$$

So, there exists k^0 s.t.

$$k > k^0 \implies \hat{\alpha}_{n_k} < \frac{1}{2}\alpha^* < 0.$$

Now, by (5)

$$L_{n_k}'(\alpha^*) \le L_{n_k}'(\hat{\alpha}_{n_k}) \qquad \forall\, k > k^0. \tag{35}$$

Recall, for $\hat{\alpha}_{n_k} < 0$ to be the MLE, we should have $L_{n_k}'(\hat{\alpha}_{n_k}) \le 0$. Then from (35) we obtain

$$L_{n_k}'\left(\frac{1}{2}\alpha^*\right) \le 0 \tag{36}$$

for all $k > k^0$ and $\frac{1}{2}\alpha^* < 0$. So for $\frac{1}{2}\alpha^* < 0$,

$$L_n'\left(\frac{1}{2}\alpha^*\right) \to -\infty \quad \text{as } n\to\infty. \tag{37}$$

Now (37), (35), and (33) imply the desired result (34).

□

Corollary 2. *Except for a P-null set of realizations, if the sequence* $\{\hat{\alpha}_n\}$ *if maximum likelihood estimates has an accumulation point* $\alpha^* \leq 0$*, then*

$$2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2 \to 0 \quad \text{as } m \to \infty \implies a_m^j \to 0, \;\; b_m^j \to 0 \quad \text{as } m \to \infty. \tag{38}$$

Recall $a(\cdot,\cdot;\cdot)$ and $b(\cdot,\cdot;\cdot)$ relate the unknown parameter α to the transition probabilities. So, (38) tells us that for $\{\hat{\alpha}_n\}$ not to converge to $\alpha^0 = 0$, the transition probabilities should give less and less information on α has time n goes on.

Theorem 13. *Under (A11), except for a P-null set of realizations,*

$$\hat{\alpha}_n \to \alpha^* \quad \text{as } n \to \infty \tag{39}$$

for $\alpha^* \in \overline{J}$*.*

From Theorem 12, we know almost surely

$$\sum_{m=0}^{\infty} \left(\sum_{j=1}^{I} \left[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2 \right] \right) = \infty \implies \hat{\alpha}_n \to \alpha^0 = 0, \tag{40}$$

so we only have to consider those realizations where

$$\sum_{m=0}^{\infty} \left(\sum_{j=1}^{I} \left[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2 \right] \right) < \infty. \tag{41}$$

Let us consider the limit behavior of $L_n'(\cdot)$ for those realizations where (41) holds. By (30)-(31) and (41) we have

$$\sum_{m=0}^{\infty} E_m(\alpha) < 0 \tag{42}$$

for $\alpha < 0$.

Consider now the following lemma, complementary in some sense to Theorem 7:

Lemma 2. *For a given $\alpha \in J$ s.t. $\alpha \neq 0$ and for every realization not in a P-null set*

$$\sum_{m=0}^{\infty} E_m(\alpha) < \infty \implies L_n'(\alpha) \to c \quad \text{as } n \to \infty. \tag{43}$$

Proof. Assume $\alpha < 0$. Recall that

$$L_n'(\alpha) = M_n + \sum_{m=0}^{n-1} E_m.$$

Since we know by Lemma 1 that $\sum_{m=0}^{n-1} E_m$ converges, we only have to prove M_n converges as $n \to \infty$. From (27) we have the following inequality

$$\sum_{m=0}^{n-1} V_m \leq \frac{1}{K|\alpha|} \sum_{m=0}^{n-1} E_m. \tag{44}$$

Convergence of $\sum_{m=0}^{n-1} E_m$ implies convergence of $\sum_{m=0}^{n-1} V_m$. Now we can use Lemma 1, which tells us that for all realizations not in a P-null set, $M_n \to M < \infty$.

□

Now, we show $L'(\alpha)$ is continuous, differentiable, and strictly decreasing. Call the limit in (43) $c = L'(\alpha)$. From (6) and (A11) we have

$$0 \geq L_n'(\alpha) \geq -\frac{1}{K^2} \sum_{m=0}^{n-1} \left[2a_m^2 + 2a_m b_m + b_m^2 - 2a_m c_m\right] \tag{45}$$

$$\geq -\frac{1}{K^2} \sum_{m=0}^{n-1} \left(\sum_{j=1}^{I} \left[\right] 2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2 \right]. \tag{46}$$

From this inequality it follows that $L_n'(\alpha)$ is a nonincreasing function converging for all α, and the limit function $c = L'(\alpha)$ is uniformly bounded in α for those realizations where (41) holds. Then

$$L''(\alpha) = -\sum_{m=0}^{n-1} \frac{2a_m^2\alpha^2 + 2a_m b_m \alpha + b_m^2 - 2a_m c_m}{(a_m\alpha^2 + b_m\alpha + c_m)^2}. \tag{47}$$

It follows that $L'(\alpha)$ is strictly decreasing for all realizations for which $\exists m$ s.t. $a_m, b_m, c_m \neq 0$. So,

Theorem 13 holds. The following lemma characterizes the behavior of $L_n'(\alpha)$:

Lemma 3. *Except for a P-null set of realizations*

$$\sum_{m=0}^{\infty}\left(\sum_{j=1}^{I}\left[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2\right]\right) < \infty \implies L_n'(\alpha) \to L'(\alpha) \quad \text{as } n \to \infty \tag{48}$$

uniformly in α.

The function $L'(\alpha)$ is continuous, differentiable, and strictly decreasing.

Let us now return to the proof of Theorem 13. We have that $\overline{J}$ is compact, so we just need to show the sequence $\{\hat{\alpha}_n\}$ does not oscillate. It is enough to show that it is not the case that $\exists [q,r] \subseteq J$ where $q < r$ and $\exists \{n_m^1\}, \{n_m^2\}$ subsequences of $1, 2, \ldots$ s.t.

$$\hat{\alpha}_{n_m^1} < q \quad \text{and} \quad \hat{\alpha}_{n_m^2} > r \qquad \forall m. \tag{49}$$

Assume (49) holds. Then from the definition of L_n and $\hat{\alpha}_n$,

$$\begin{aligned} &L_{n_m^1}(q) \le 0, \qquad L_{n_m^1}(r) \le 0 \qquad &\forall\, m \\ &L_{n_m^2}(q) \ge 0, \qquad L_{n_m^2}(r) \ge 0 \qquad &\forall\, m, \end{aligned}$$

so as the limits exist by Lemma 3, it follows that $L'(q) = L'(r) = 0$, but this cannot be the case as $q \neq r$ and $L'(\cdot)$ is strictly decreasing.

5.3 Adaptive Control and Results

Having the convergence of estimates of our unknown parameter, we can now use adaptive control. We will make another assumption:

(A14) Suppose

$$u_n = \phi(\alpha_n, X_n)$$

and $a(i,j;\phi(\alpha,i))$ and $b(i,j;\phi(\alpha,i))$ are continuous functions in α for all $i,j \in I$.

Proposition 12. *Assume* (A11) *and* (A14). *Except for a P-null set, if the sequence* $\{\hat{\alpha}_n\} \to \alpha^* \neq$

$\alpha^\circ = 0$ *as* $n \to \infty$, *then*

$$a(i, j; \phi(\alpha^*, i)) = 0, \qquad b(i, j; \phi(\alpha^*, i)) = 0 \tag{50}$$

for every $j \in I$ *and for every state* i *that is reached i.o.*

Proof. If we discard the P-null set where (39) does not hold, then we have that $\hat{\alpha}_n \to \alpha^*$ as $n \to \infty$. By Corollary 2,

$$0 = \lim_{m\to\infty} a_m^j = \lim_{m\to\infty} a(X_m, j; \phi(\hat{\alpha}_m, X_m)), \quad 0 = \lim_{m\to\infty} b_m^j = \lim_{m\to\infty} b(X_m, j; \phi(\hat{\alpha}_m, X_m)) \tag{51}$$

for $j = 1, \ldots, I$. If a state i is reached i.o. we can take a subsequence m_k s.t. $X_{m_k} = i$ for all k. Along this sequence,

$$\begin{aligned} 0 &= \lim_{k\to\infty} a(X_{m_k}, j; \phi(\hat{\alpha}_{m_k}, X_{m_k})) = \lim_{k\to\infty} a(i, j; \phi(\hat{\alpha}_{m_k}, i)) = a(i, j; \phi(\alpha^*, i)), \\ 0 &= \lim_{k\to\infty} b(X_{m_k}, j; \phi(\hat{\alpha}_{m_k}, X_{m_k})) = \lim_{k\to\infty} b(i, j; \phi(\hat{\alpha}_{m_k}, i)) = b(i, j; \phi(\alpha^*, i)), \end{aligned}$$

where we have used the continuity from (A14).

□

We would like to extend (51) to all states in I, so we will make the following assumption:

(A15) For all $i, j \in I$ there exists the sequence $i_0, \ldots, i_r$ s.t.

$$p(i_{s-1}, i_s; u, \alpha) > 0$$

for $s = 1, \ldots, r + 1$ where $i_0 = i$ and $i_r = j$.

From Theorem 11, it follows under (A15) that all states are reached i.o. almost surely, and thus Proposition 12 holds for all $i \in I$. Now, we also have the following proposition:

Proposition 13. *Assume* (A11), (A14), (A15). *Except for a* P*-null set of realizations, if the*

sequence $\{\hat{\alpha}_n\}$ *converges to* α^**, then*

$$p(i,j;\phi(\alpha^*,i),\alpha^*) = p(i,j;\phi(\alpha^*,i),\alpha^\circ) \tag{52}$$

for every $i,j \in I$*.*

Proof. The equation (52) holds if $\alpha^* = \alpha^0$. If $\alpha^* \neq \alpha^\circ$, let us consider the set of all realizations where Proposition 12 holds and every state i is reached i.o. Then

$$\begin{aligned}
&p(i,j;\phi(\alpha^*,i),\alpha^*) - p(i,j;\phi(\alpha^*,i),\alpha^\circ) \\
&\qquad = a(i,j;\phi(\alpha^*,i))(\alpha^*)^2 + b(i,j;\phi(\alpha^*,i))\alpha^* + c(i,j;\phi(\alpha^*,i)) \\
&\qquad\quad - a(i,j;\phi(\alpha^*,i))(\alpha^\circ)^2 - b(i,j;\phi(\alpha^*,i))\alpha^\circ - c(i,j;\phi(\alpha^*,i)) \\
&\qquad = a(i,j;\phi(\alpha^*,i))\big[(\alpha^*)^2 - (\alpha^\circ)^2\big] + b(i,j;\phi(\alpha^*,i))\big[\alpha^* - \alpha^\circ\big] + 0 \\
&\qquad = 0 \cdot \big[(\alpha^*)^2 - (\alpha^\circ)^2\big] + 0 \cdot \big[\alpha^* - \alpha^\circ\big] \qquad \text{(by Proposition 12)} \\
&\qquad = 0.
\end{aligned}$$

As in Theorem 7 in [1], we can consider $\phi(\alpha,\cdot)$ as giving a feedback control law to be used when we know α, and (52) can be interpreted as, under the control law specified by the estimated value α^*, α° is indistinguishable from α^*. Theorem 8 provides a rule to modify the control law $\phi(\cdot,\cdot)$ to guarantee convergence to the true value α°: If necessary, modify $\phi(\alpha_n, X_n)$ to have

$$\sum_{m=0}^{\infty}\left(\sum_{j=1}^{I}\big[2(a_m^j)^2 - 3a_m^j b_m^j + (b_m^j)^2\big]\right) = \infty.$$

□

5.4 Concluding Remarks and Future Investigations

We have now shown, in support of previous results, that the maximum likelihood estimate converges to a value α^* indistinguishable from the true value under a control feedback law induced by α^*. More specifically, we have shown that all results by Sagalovsky in [8] extend to the quadratic case with the additional assumptions discussed.

As mentioned in [7], these results can be extended to controlled Markov chains with transition probabilities depending on α by higher degree polynomials, but the computations become very difficult. Additionally, no known effort has been made to consider a controlled Markov chain with transition probabilities depending on α via a convex function. Thus, there should be future investigations on such a model to contribute to this collection of results. Furthermore, computational investigations of these results would add a new dimension and possibly lead to further interest based on useful application of adaptive control of Markov chains. Lastly, most of the above results use maximum likelihood estimation to estimate the parameter α. Investigation into other forms of estimation, notably weighted least squares estimation, and a behavioral comparison to the asymptotic behavior of the maximum likelihood estimation would be interesting. Further suggestions for continued investigations can be found in [1], [3], [2], [7], [8].

References

[1] V. Borkar and P. Varaiya. Adaptive Control of Markov Chains, I: Finite Parameter Set. *IEEE Trans. on Autom. Control*, 24(6): 953-957, Dec 1979.

[2] P.R. Kumar. A Survey of Some Results in Stochastic Adaptive Control. *SIAM J. on Control and Optim.*, 23(3): 329-380, 1985.

[3] P.R. Kumar. Adaptive Control with a Compact Parameter Set. *SIAM J. of Control and Optim.*, 20(1): 9-13, Jan 1982.

[4] P.R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice-Hall, Inc., New Jersey, 1986.

[5] P. Mandl. A Connection Between Controlled Markov Chains and Martingales. *Kybernetika*, 9(4): 237-241, 1973.

[6] P. Mandl. Estimation and Control in Markov Chains. *Adv. in Applied Prob.*, 6(1): 40-60, Mar 1974.

[7] B. Pasik-Duncan. *On Adaptive Control*. SGPiS-Publishers, Warsaw, 1986.

[8] B. Sagalovsky. Adaptive Control and Parameter Estimation in Markov Chains: A Linear Case. *IEEE Trans. on Autom. Control*, 27(2): 414-419, Apr 1982.