UC Irvine
UC Irvine Electronic Theses and Dissertations

Title

Absorbing Markov Chains with Random Transition Matrices and Applications

Permalink

Author

Kassir, Ali

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, IRVINE

Absorbing Markov Chains with Random Transition Matrices and Applications

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Ali Kassir

Dissertation Committee:

Professor Patrick Guidotti, Co-Chair

Professor Knut Solna, Co-Chair

Professor Natalia Komarova

2018

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to acknowledge my advisors, Dr. Guidotti and Dr. Solna for their guidance, wisdom, support and patience. They believed in me before I did and have left an everlasting impact on me that I will forever be grateful for. I would like to acknowledge Dr. Komarova for her availability and kind heart. She was of extreme help at a time of great need. I would like to acknowledge my colleague Floyd Maseda for his interest in hearing what I had to share and indirectly helping me develop some of my most significant thesis results. I would like to acknowledge my parents, Hassan Kassir and Ghada Kassir, for instilling in me the passion for the pursuit and development of knowledge. I would like to acknowledge my wife, Zeinab Kachakeche, for her support and sacrifice. I hope to one day give her the same support that I received from her and to sacrifice for her as much as she has sacrificed for me. Last but not least, I would like to acknowledge my children, Hassan Kassir and Fatima Kassir, for their smiles, cries, laughter and love. I hope to one day see them reaching the highest levels of education and success.

Curriculum Vitae

Ali Kassir

Education

2013-2018	PhD in Mathematics, University of California, Irvine
2011-2013	Masters in Mathematics, University of California, Irvine
2009-2011	B.S. in Mathematics, University of California, Irvine

Abstract of the Dissertation

Absorbing Markov Chains with Random Transition Matrices and Applications

by

Ali Kassir

Doctor of Philosophy in Mathematics

University of California, Irvine, 2018

Professor Patrick Guidotti, Co-Chair

Professor Knut Solna, Co-Chair

In this thesis, we explore Markov chains with random transition matrices. Such chains are a development on classic Markov chains where the transition matrix is taken to be random. The intuition for this is that we may be interested in modeling phenomenas where the homogeneity assumption of classic Markov chains is invalid. We first use such chains to model credit risk. The randomness of the transition matrix is used to represent the randomness of the economy that underlies credit risk. With this, we model a portfolio of loans and the risk due to having a shared economy. We then proceed to explore theoretical properties of such chains with a focus on their asymptotic behavior. In the case of absorbing chains, we show that the the infinite product of independent and identically distributed random matrices must converge almost surely. We also introduce perturbed Markov chains

as a special form of Markov chains with random transition matrices. Perturbed Markov chains are Markov chains with transition matrices of the form $P + \varepsilon Q$ where P is taken to be a constant matrix, $\varepsilon \in (0, 1)$ a constant and Q random with rows summing to 0. For perturbed Markov chains with P irreducible, we approximate the long-run fluctuation of such chains up to order ε^2. As for perturbed Markov chains with P having absorbing states, we approximate the variance of the fundamental matrix and mean time to absorption. We also explore some applications. In the Moran process application, we study the impact of temporal randomness on the Moran process and derive an analytic result to calculate the probability of fixation.

Chapter 1

Introduction

Markov chains are among the most well known and established probabilistic models. They have long been developed, studied and applied to real world problems. In this thesis we are particularly interested in discrete-time Markov chains. The simplicity, yet richness of these models makes them interesting in theoretical studies as well as extremely versatile in applications. A Markov chain is a stochastic process $\{X_t\}$ indexed by time on a state space $S = \{1, \ldots, s\}$ such that

$$P\left(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0\right) = P\left(X_{t+1} = j \mid X_t = i\right).$$

Abiding by the convention $p_{t,ij} = P\left(X_{t+1} = j \mid X_t = i\right)$ gives rise to the $s \times s$ (time dependent) transition matrix $P_t = [p_{t,ij}]$. Classic Markov chain models come in two forms, homogeneous and inhomogeneous. If we assume that the transition matrix is independent of time, then we will have a homogeneous Markov chain. Otherwise, the Markov chain is inhomogeneous. Most of the study that pertains to discrete-time Markov chains is in regards to homogeneous Markov chains. As a result, we find that many Markov chain applications impose this assumption, even though it might not be entirely accurate due to a natural level of environmental randomness that takes place. Our main focus in this thesis will be centered around this point: to explore a framework that allows for randomness in the transition

matrix P.

In the first portion of this thesis, we will explore a Markov chain application to credit risk. We develop on a classic Markov model that was initially proposed by Jarrow et al. in [21]. While the model that Jarrow et al. present appropriately models some aspects of credit risk, it has its short comings. Our research stems from credit risk migrations failing to satisfy the homogeneity assumption in their Markov model. This is discussed in [25, 4, 18, 16, 14]. In our model, an important element is to capture the dependence of the transition matrix on the state of the economy. This dependence can be interpreted as a form of a correlation factor that is taking place between different entities in the same economy. We will introduce randomness into the problem by assuming that the transition matrix itself is random and may change based on the economy. By doing so, we are able to capture the economic impact of credit risk on an entire portfolio.

We then proceed beyond this application to explore the theory of Markov chains for which the transition matrix is random. Such Markov chains have been explored in the literature and have been referred to as Markov chains in random environments in [11, 9, 10, 12, 13, 6, 34, 35, 3, 2, 31, 32, 37]. In a 1969 paper, Takahashi [33] presents his formulation of Markov chains with random transition matrices and generalizes the concepts of stationarity, irreducibility, ergodicity and periodicity. He also proves several ergodic theorems. In a more recent sequence of papers [27, 28, 29, 30], Saloff-Coste and Zungia study the asymptotic behavior of time inhomogeneous Markov chains as well as develop the notions of stability and merging. Independently of the above authors, Brueneau et al. [7] study infinite products of random transition matrices within the more general context of dynamical systems and applications to physics. They also arrive at some similar results in regards to the behavior of such products. In particular, they show that the product of random transition matrices can be decomposed into a sum of a fluctuating and an exponentially decaying process, where the fluctuating part converges in Cesàro mean. The results presented by Takahashi and Bruneau are limited to matrices with some form of irreducibility. In particular they don't consider

nor discuss absorbing Markov chains with random transition matrices.

In a sequence of papers [11, 9, 10, 12, 13, 6] published between 1980 and 1991, Cogburn explores Markov chains in random environments in the case that the environmental randomness is Markovian. This extra assumption creates a structure that is different from that presented by Takahashi and Brueneau et al., where they take the matrices in the different time steps to be independent and identically distributed. In [11], Markov chains in Markovian environments are established and then a classification of the state space of the environment is presented. In [6], Bourgin and Cogburn present results on the absorption probabilities for such Markov chains. In [9, 10, 12], results about the asymptotic behavior of random products of stochastic matrices are presented. In [34, 35], Torrez also studies birth and death processes with such a Markov structure and arrives at results pertaining to the extinction of such processes. In [2, 3], Athreya and Karlin study branching processes in random environments where the randomness is independent and identically distributed with respect to time. Smith and Wilkinson also study branching process with such a structure in [31, 32, 37].

Our work primarily develops on the framework that is established by Bueneau et al. and Takahashi. We prove that an absorbing Markov chain with random transition matrices that share absorbing states converges almost surely. Let $P(\omega)$ be a random matrix that takes values from $\mathcal{P}^{(r)}$, the set of transition matrices that share the first r states as absorbing states. Then then the infinite product of independent and identically distributed matrices $\prod_{i=1}^{\infty} P(\omega_i)$ converges almost surely. We also, present perturbed Markov chains. The reason for establishing and studying perturbed Markov chains is because we are interested in studying Markov chains that are "almost" homogeneous. In particular, we want to study the behavior of inhomogeneous Markov chains that are driven by a sequence of random perturbations of a transition matrix P. To illustrate, consider the following formulation. Let $\mathcal{Q}$ denote the set of $s \times s$ matrices with rows summing to 0 and let $Q : \Omega \to \mathcal{Q}$ be a random matrix on the probability space $(\Omega, \mathcal{F}, \mathrm{Pr})$. Moreover, assume $\mathbb{E}[Q] = \mathbb{O}$, the zero matrix.

For $\varepsilon \in (0, 1)$, let $\mathcal{P} = \{P + \varepsilon Q(\omega) \mid \omega \in \Omega\}$ be the set of perturbed transition matrices. We want to study inhomogeneous Markov chains driven by $\mathcal{P}$. The transition matrix of such a chain over an n-step time period is a random matrix $P^{(n)} : \Omega^n \to [0, 1]^{s \times s}$ given by

$$P^{(n)}(\omega_1, \ldots, \omega_n) = \prod_{j=1}^{n} (P + \varepsilon Q(\omega_j)).$$

We explore the randomness of $P^{(n)}$, as well as associated quantities such as $\mathbb{E}\left[P^{(n)}\right]$, $\mathrm{Var}\left(P_{ij}^{(n)}\right)$ and $\mathrm{Cov}\left(P_{ij}^{(n)}, P_{h\ell}^{(n+k)}\right)$ for fixed $k \geq 0$ in the large n limit. In Proposition 4.3.1, we present an analytic approximation for the covariance of the entries of $P^{(n)}$ in the limit. After establishing the convergence of absorbing Markov chains with random transition matrices, we look at the time until absorption and come up with an analytic approximation for the variance of the entries of the fundamental matrix in Proposition 5.3.1.

Upon establishing the theoretical foundation for absorbing Markov chains with random transition matrices and their convergence, we explore two applications. In the first application, we present and study a variation of the gambler's ruin problem. We explore the asymptotic behavior of the probabilities of transition as well as the time to absorption. In our second application, we present a formulation of a Moran process with temporal randomness. A Moran process is a classic Markov chain model with a transition matrix that takes a special form. Moran processes are is used in modeling biological systems. We extend this model by incorporating randomness in the transition matrices as a form of time-dependent randomness. We then study how this temporal randomness impacts the probability of fixation and time to absorption of the process. In Sections 6.2.4 and 6.2.5, we derive an analytic expression for the probability of fixation.

The rest of this thesis will be organized in the following way. In Chapter 2, we will present an overview of the classic Markov chain results that pertain to our work. In Chapter 3, we will present an economy switching Markov model that takes into account the economy as a correlation factor. In Chapter 4, we give an overview of Markov chains with random

transition matrices, some available results and establish perturbed Markov chains. In Chapter 5, we establish some results in regards to the asymptotic behavior of absorbing Markov chains with random transition matrices. Finally in Chapter 6, we present some applications of our model to the gambler's ruin problem and to Moran processes.

Chapter 2

An Overview of Markov Chains

In this chapter, we will be presenting an overview of the classic Markov chain results that pertain to our work. We will discuss basic definitions and theorems on absorbing and ergodic Markov chains. For a more detailed discussion, we refer the reader to [19, 20, 22].

2.1   Introduction

A Markov chain is a random process that changes at discrete integer-valued time intervals with the property that only the current value of the chain determines its distribution at the next time step. At each integer time step $n = 0, 1, \ldots$, there is an integer valued random variable X_n taking on values from a state space S. The state space S can either be countably infinite, in which case it is usually taken to be $S = \{0, 1, 2, \ldots\}$, or it can be finite, in which case it can be taken to be of the form $S = \{1, \ldots, s\}$ for some s. Markov chains study the random process that starts in one of these states and moves successively from one state to another. In our work, we will limit our discussion to finite-state Markov chains. A Markov chain can be viewed as a family of random variables $\{X_n \mid n \geq 0\}$.

Definition 2.1.1. A *Markov chain* is a family of random variables $\{X_n \mid n \geq 0\}$ indexed by discrete-time where each X_n is an S-valued random variable for some finite state space $S = \{1, \ldots, s\}$ and satisfies the following Markov property: for all positive integers n, and for all choices of $i_0, \ldots, i_{n-2}, i, j \in S$,

$$\Pr\left(X_n = j \mid X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0\right) = \Pr\left(X_n = j \mid X_{n-1} = i\right). \qquad (2.1.1)$$

Furthermore, we will call the Markov chain *homogenous* in the case that $\Pr\left(X_n = j \mid X_{n-1} = i\right)$ does not depend on n and the values $P_{ij} = \Pr\left(X_n = j \mid X_{n-1} = i\right)$ can be written in matrix form as the (i,j)-th entries of the $s \times s$ transition (or migration) probability matrix P. In the case that $P_{ij}(n)$ depends on n, we will call the Markov chain *inhomogeneous*.

Equation 2.1.1 describes the important characteristic of Markov chains that the current state summarizes everything about the past and is the only relevant information for the future. The initial probability distribution for X_0 along with the transition matrix P define the probabilities for all possible events in the Markov chain. Homogeneous Markov chains will be the main focus of this chapter and from here on homogeneous Markov chains will be simply referred to as Markov chains. Inhomogeneous Markov chains will not be studied in this chapter, but will play an interesting role in the next chapters.

The matrix representation of the Markov chain gives rise to several important properties of P. Let P_{ij}^n denote the (i,j)-th entry of P^n, the n-th power of P.

Theorem 2.1.1. *Let $\{X_n\}$ be a Markov chain with state space $S = \{1, \ldots, s\}$ and transition probabilities matrix P. Then for all $i, j \in S$ and positive integers n, m,*

1. $\displaystyle\sum_{k=1}^{s} P_{ik} = 1$

2. $Pr\left(X_n = j \mid X_0 = i\right) = P_{ij}^n$

3. *Chapman-Kolmogorov Equation:* $P_{ij}^{m+n} = \sum_{k=1}^{s} P_{ik}^m P_{kj}^n$

Proof.

1. This is due to the fact that each row of P represents the probability mass function of the Markov chain at the next time step, given current state i.

2. Using the Markov property 2.1.1, we have

$$\Pr\left(X_2 = j \mid X_0 = i\right) = \sum_{k=1}^{s} \Pr\left(X_2 = j \mid X_1 = k\right) \Pr\left(X_1 = k \mid X_0 = i\right)$$
$$= \sum_{k=1}^{s} P_{ik} P_{kj} = P_{ij}^2.$$

By induction $\Pr\left(X_n = j \mid X_0 = i\right) = P_{ij}^n$ will hold for all n.

3. This equation is a direct consequence of the above.

$\square$

As a direct consequence of this theorem, if we know the initial distribution of a Markov chain, then we can calculate the chain's distribution at each time step.

Corollary 2.1.1. *Let $\alpha^{(0)}$ be a probability vector representing the probability mass function of the initial state X_0 of a Markov chain with transition matrix P. Then the probability that the chain is in state i after n steps is the i-th entry of the vector*

$$\alpha^{(n)} = \alpha^{(0)} P^n.$$

The possibility of a Markov chain to migrate from a given state to another state over several time steps, or the lack thereof, plays an important role in its long term behavior. For this purpose, it is important to present the following definitions for the classification of the states.

Definition 2.1.2. Let P be a transition matrix for a Markov chain.

1. A state j is *accessible* from i if there exists n such that $P_{ij}^n > 0$.

2. Two distinct states i and j *communicate* if i is accessible from j and j is accessible from i.

3. A *recurrent* state is a state i that is accessible from all states that are accessible from i.

4. A *transient* state is a state that is not recurrent.

Note that we can decompose the state space of a Markov chain into disjoint subsets comprised of the states that communicate with each other and the states that don't. A set of states that communicate with each other will form a class C in the sense that for $i, j \in S$, i and j communicate with each other if and only if $i, j \in C$. Moreover, either all states in a class are transient or all are recurrent.

Another important classification of states is with respect to their periods.

Definition 2.1.3. The *period of a state* i, denoted as $d(i)$ is defined with respect to the greatest common divisor (gcd) as

$$d(i) := gcd(n \mid P_{ii}^n > 0).$$

A state is *aperiodic* if is has period 1 and is *periodic* otherwise.

A simple yet import fact is that for any Markov chain, all states in the same class have the same period. For a proof visit [19].

2.2 Absorbing Markov Chains

One special type of Markov chains that we will be working with is called an absorbing Markov chain.

Definition 2.2.1. A state i of a Markov chain is called *absorbing* if it is impossible to leave it, that is $P_{ii} = 1$. A Markov chain is *absorbing* if it has at least one absorbing state that is

accessible from every other state.

Note that in an absorbing Markov chain, all the non-absorbing states are transient. Consider an absorbing chain with r absorbing states and t transient states. Renumbering the states so that the transient states come first, we can write the transition matrix in the canonical form

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

where Q is $t \times t$, R is $t \times r$ and I is the $r \times r$ identity matrix. Due to the special form of P, the powers will be

$$P^n = \begin{bmatrix} Q^n & (I + Q + \cdots + Q^{n-1}) R \\ 0 & I \end{bmatrix}.$$

This form of P^n shows that the entries of Q^n give the probabilities of remaining in the transient states. In fact, we see in the following theorem that $Q^n \to 0$, the zero matrix.

Theorem 2.2.1. *With probability 1, the process with transition matrix*

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

will be absorbed.

Proof. Let m_j be the minimum number of steps required to reach an absorbing state from state j. Let p_j be the probability that, starting from state j, the process will not be absorbed in m_j steps. Then $p_j < 1$. Let $m = \max_j m_j$ and $p = \max_j p_j$. The probability of not being absorbed in km steps is less than or equal to p^k, which must tend to 0 since $p < 1$. Since the probability of not being absorbed in n steps is monotone decreasing, these probabilities also tend to 0, hence $\lim_{n \to \infty} Q^n = 0$. $\qquad \square$

Theorem 2.2.2. *The inverse of $I - Q$ exists, call it N. Moreover, for $i, j \leq t$, the entry N_{ij} of N is the expected number of visits to state j given initial state i.*

Proof. To show that $(I - Q)^{-1}$ exists, it suffices to show that if $(I - Q)\,x = 0$ then $x = 0$. If $(I - Q)\,x = 0$ then $x = Qx$ and hence iterating we have $x = Q^n x$. Since $Q^n \to 0$, we $x = 0$.

Let $N = (I - Q)^{-1}$. Then $N = I + Q + Q^2 + \cdots$. Indeed,

$$(I - Q)\left(I + Q + Q^2 + \cdots + Q^n\right) = I - Q^{n+1}$$

and hence

$$I + Q + Q^2 + \cdots + Q^n = N\left(I - Q^{n+1}\right).$$

Letting $n \to \infty$ we have $N = I + Q + Q^2 + \cdots$.

Fix i and j with $i, j \leq t$. Let $X_{ij}^{(\ell)}$ be a random variable which equals 1 if the chain, with initial state i, is in state j after ℓ steps, and equals 0 otherwise. Then

$$P\left(X_{ij}^{(\ell)} = 1\right) = Q_{ij}^{(\ell)} \qquad \text{and} \qquad P\left(X_{ij}^{(\ell)} = 0\right) = 1 - Q_{ij}^{(\ell)},$$

where $Q_{ij}^{(\ell)}$ is the ij-th entry of Q^ℓ, with the notation $Q^0 = I$. Then

$$E\left[X_{ij}^{(\ell)}\right] = 1 \cdot Q_{ij}^{(\ell)} + 0 \cdot \left(1 - Q_{ij}^{(\ell)}\right) = Q_{ij}^{(\ell)}$$

and hence the expected number of times the chain is in state j in the first n steps is

$$E\left[\sum_{k=0}^{n} X_{ij}^{(k)}\right] = \sum_{k=0}^{n} Q_{ij}^{(k)}.$$

Let $X_{ij} = X_{ij}^{(0)} + X_{ij}^{(1)} + \cdots$. Note that X_{ij} gives the total number of times the chain is in state j before absorption. Moreover, X_{ij} must be finite by the previous theorem, with probability 1. Therefore,

$$N_{ij} = E\left[X_{ij}\right] = \sum_{k=0}^{\infty} Q_{ij}^{(k)}.$$

$\square$

11

Recall the basic fact about conditional expectations that if $\{B_k\}$ partitions the sample space, then

$$E\left[f\left(X\right)\right] = \sum E\left[f\left(X\right) \mid B_k\right] P\left(B_k\right).$$

The following is another proof of the second half of Theorem 2 that provides another perspective and some intuition.

Proof. As above, let X_{ij} be the total number of times the chain is in state j given initial state i. Note that $X_{ij} = \delta_{ij} + X_{kj}$, where δ_{ij} is the contribution of the initial state, that is, we only have a contribution if the initial state is the state that we want to calculate the expected stay in, and X_{kj} is the contribution after the first migration to some state k. Also, $X_{\ell j} = 0$ for ℓ absorbing since once the process has been absorbed, it can no longer reach S_j.

Let B_k be the event that the initial migration is to state S_k for $k = 1, \ldots, s$. Then

$$
\begin{aligned}
E\left[X_{ij}\right] &= \sum_{k=1}^{s} P_{ik} E\left[X_{ij} \mid B_k\right] \\
&= \sum_{k=1}^{s} P_{ik} E\left[\delta_{ij} + X_{kj}\right] \\
&= \sum_{k=1}^{s} P_{ik}\delta_{ij} + P_{ik} E\left[X_{kj}\right] \\
&= \delta_{ij} + \sum_{k=1}^{t} P_{ik} E\left[X_{kj}\right].
\end{aligned}
$$

Hence, re-writing in matrix notation we have $[E\left[X_{ij}\right]] = I + Q \cdot [E\left[X_{ij}\right]]$, where $[A_{ij}]$ denotes the matrix with entries A_{ij}. Therefore,

$$[E\left[X_{ij}\right]] = (I - Q)^{-1}.$$

□

Definition 2.2.2. The matrix $N = (I - Q)^{-1}$ is called the *fundamental matrix* for P.

Due to the above results, it follows that for

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix},$$

we have

$$\lim_{n \to \infty} P^n = \begin{bmatrix} 0 & NR \\ 0 & I \end{bmatrix}.$$

Theorem 2.2.3. *Let n_i be the expected number of steps before absorption, given initial state i, and let n be the column vector whose i-th entry is n_i. Then*

$$n = N\mathbf{1},$$

where $\mathbf{1}$ is the column vector all of whose entries are 1.

Proof. Note that $N\mathbf{1}$ will be a column vector with t entries. Let Z_i denote the number of steps before default, given initial state i. Then

$$n_i = E\left[Z_i\right] = E\left[\sum_{j=1}^{t} X_{ij}\right] = \sum_{j=1}^{t} n_{ij}.$$

$\square$

Theorem 2.2.4. *Let M_{ij} be the variance on the number of visits to state j given initial state i and let M be the matrix with entries M_{ij}. Then*

$$M = N\left(2N_{dg} - I\right) - N_{sq},$$

where N_{dg} is the diagonal matrix with same diagonal as N and N_{sq} is the Hadamard product of N with itself (i.e. each entry of N is squared).

Proof. Using the same notation as above, let X_{ij} be the total number of times the chain is

13

in state j given initial state i. Then

$$M_{ij} := \text{Var}\,[X_{ij}] = E\,[X_{ij}^2] - (E\,[X_{ij}])^2\,.$$

By Theorem 2.2.2, $[E\,[X_{ij}]^2] = N_{sq}$. It suffices to show that $[E\,[X_{ij}^2]] = N\,(2N_{dg} - I)$. Note,

$$
\begin{aligned}
E\,[X_{ij}^2] &= \sum_{k=1}^{s} P_{ik} E\,[X_{ij}^2 \mid B_k] \\
&= \sum_{k=1}^{s} P_{ik} E\,[(\delta_{ij} + X_{kj})^2] \\
&= \sum_{k=1}^{s} P_{ik} E\,[\delta_{ij}^2 + 2X_{kj}\delta_{ij} + X_{kj}^2] \\
&= \sum_{k=1}^{s} P_{ik}\,(\delta_{ij} + 2E\,[X_{kj}]\,\delta_{ij} + E\,[X_{kj}^2]) \\
&= \delta_{ij} + \sum_{k=1}^{t}\left(2P_{ik}E\,[X_{kj}]\,\delta_{ij} + P_{ik}E\,[X_{kj}^2]\right).
\end{aligned}
$$

Thus,

$$[E\,[X_{ij}^2]] = I + 2\,(QN)_{dg} + Q\cdot[E\,[X_{ij}^2]]$$

and therefore,

$$
\begin{aligned}
[E\,[X_{ij}^2]] &= (I - Q)^{-1}\left(2\,(QN)_{dg} + I\right) \\
&= N\left(2\,(N - I)_{dg} + I\right) \\
&= N\,(2N_{dg} - I),
\end{aligned}
$$

where the second equality holds since $N - I = Q + Q^2 + \cdots = QN$. $\qquad\square$

Theorem 2.2.5. *Let r_i be the variance on the number of steps before absorption with initial*

state i, and let r be the column vector whose i-th entry is r_i. Then

$$r = (2N - I)\, n - n_{sq},$$

where n_{sq} is the Hadamard product of n with itself.

Proof. As in the proofs above, let Z_i denote the number of steps before absorption, given initial state i, and note that $Z_\ell = 0$ for absorbing states ℓ. Then

$$r_i := \operatorname{Var}[Z_i] = E\left[Z_i^2\right] - E\left[Z_i\right]^2.$$

Note,

$$
\begin{aligned}
E\left[Z_i^2\right] &= \sum_{k=1}^{s} E\left[Z_i^2 \mid B_k\right] P(B_k) \\
&= \sum_{k=1}^{s} P_{ik} E\left[(1 + Z_k)^2\right] \\
&= \sum_{k=1}^{s} P_{ik} E\left[1 + 2Z_k + Z_k^2\right] \\
&= \sum_{k=1}^{s} P_{ik}\left(1 + 2E\left[Z_k\right] + E\left[Z_k^2\right]\right) \\
&= 1 + \sum_{k=1}^{t} 2P_{ik} E\left[Z_k\right] + P_{ik} E\left[Z_k^2\right].
\end{aligned}
$$

Thus,

$$\left[E\left[Z_i^2\right]\right] = \mathbf{1} + 2Qn + Q \cdot \left[E\left[Z_i^2\right]\right]$$

15

and hence

$$
\begin{aligned}
\left[E\left[Z_i^2\right]\right] &= (I-Q)^{-1}\left(\mathbf{1}+2Qn\right) \\
&= N\mathbf{1}+2NQn \\
&= t+2\left(N-I\right)n \\
&= \left(2N-I\right)n.
\end{aligned}
$$

Therefore,

$$
\left[\mathrm{Var}\left[Z_i\right]\right] = \left(2N-I\right)n - n_{sq},
$$

since $\left[E\left[Z_i\right]^2\right] = n_{sq}$ by Theorem 2.2.3. $\qquad\square$

Theorem 2.2.6. *Let H_{ij} be the probability of visiting state j when starting with initial state i, and let H be the matrix with entries H_{ij}. Then*

$$
H = \left(N-I\right)N_{dg}^{-1},
$$

where N_{dg} is the diagonal matrix with same diagonal as N.

Proof. Let B be the event that the process reverts to state j. Note that $\{X_{ij}\mid B\} = \{\delta_{ij}+X_{jj}\}$ as events. Indeed, the RHS can be interpreted as the sum of the contribution of the initial state and the contribution beginning with the first migration to state j and beyond. Also, $\{X_{ij}\mid B^c\} = \delta_{ij}$ because the only contribution will be from the initial state. So,

$$
\begin{aligned}
E\left[X_{ij}\right] &= E\left[X_{ij}\mid B\right]P\left(B\right)+E\left[X_{ij}\mid B^c\right]P\left(B^c\right) \\
&= E\left[\delta_{ij}+X_{jj}\right]H_{ij}+\delta_{ij}\left(1-H_{ij}\right) \\
&= \left(\delta_{ij}+E\left[X_{jj}\right]\right)H_{ij}+\delta_{ij}\left(1-H_{ij}\right) \\
&= E\left[X_{jj}\right]H_{ij}+\delta_{ij}.
\end{aligned}
$$

16

Thus, $N = HN_{dg} + I$ and hence

$$H = (N - I) N_{dg}^{-1}.$$

□

2.3 Ergodic Markov Chains

Another family of Markov chains that is relevant to our work is ergodic Markov chains. Consider the following definitions.

Definition 2.3.1. A Markov chain is called a *unichain* if it contains a single recurrent class (and possibly some transient states). A Markov chain is *ergodic* if it consists of only one class of states that is both recurrent and aperiodic. An *ergodic unichain* is a unichain for which the recurrent class is ergodic.

With these definitions established, we want to discuss the asymptotic behavior of P^n. In the case of ergodic unichains, the long term behavior of the chain will be independent of the initial states. That is $P_{ij}^n \to v_j$ and hence $P^n \to \mathbf{1}v$ where v is the row vector with entries v_i and is known as the invariant distribution. Moreover, the invariant distribution will have the special property that $vP = v$. An important theorem is the following:

Theorem 2.3.1. *The powers of the transition matrix P converge to a rank 1 matrix of the form $\mathbf{1}v$ if and only if P is the transition matrix of an ergodic unichain. If P has several recurrent classes, then P^n converges to a matrix with non-identical rows.*

We will discuss a sketch of the proof, focusing on the main idea behind the proof and the role that eigenvalues and the Jordan decomposition play. For a more rigorous proof, we direct the reader to [19].

For any transition matrix P, its eigenvalues must have magnitude at most 1. In the case P is the transition matrix of an ergodic unichain having all distinct eigenvalues $1 = \lambda_1 >$

17

$\lambda_2 > \cdots > \lambda_s$, the Jordan decomposition of P will be of the form

$$P = U\Lambda U^{-1}$$

for some U where Λ is the diagonal matrix with entries λ_i. So

$$P^n = U\Lambda^n U^{-1} = U\Lambda_1 U^{-1} + U\left(\Lambda^n - \Lambda_1\right) U^{-1}$$

where

$$\Lambda_1 = \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \quad \text{and} \quad \Lambda^n - \Lambda_1 = \begin{bmatrix} 0 & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_s^n \end{bmatrix}.$$

Thus, P^n will converge geometrically to $U\Lambda_1 U^{-1}$ which will be a matrix with identical rows.

In the case that P is the transition matrix of a chain with several recurrent classes, the Jordan form of P will have several blocks $B_1, \ldots, B_\ell$. The Jordan blocks associated with the eigenvalue 1 will determine the convergence, while the other components will tend to 0.

2.4 Looking Forward

The discussion presented in this chapter lays out some of the background material that is relevant to the content of this thesis. The material presented in Section 2.2 will be important in our economy switching Markov model presented in Chapter 3 and the applications content in Chapter 6. The intuition that underlies Theorem 2.3.1 will be important in some of the results of Chapters 4 and 5.

Chapter 3

Economy Switching Markov Model

Since the late 1990s, Markov chains have been used to model credit risk because of their ability to capture the natural credit rating migration behavior that takes place by entities in the credit market. However, one issue with the basic model that is agreed upon in the literature is that the homogeneity assumption of the transition matrix is too strong [25, 8, 14, 16, 17, 18, 4]. In this chapter, we will begin by presenting Jarrow et al.'s Markov chain model of credit risk. We will then develop this model, in particular modifying the homogeneity assumption. Our approach is an economy switching Markov model. It involves considering two transition matrices, one associated with healthy economic times and another associated with difficult economic times, and a small probability parameter that allows the process to switch between the two matrices. This approach is intended to capture the shared impact that the economy has on the different entities due to a shared economy. Taking this into account will allow us to model the higher risk that is associated with economic recessions. With this model, we will present results in regards to the distribution of the number of collected payments of a portfolio of loans. We will also run a simulation and calculate the associated distributions, value-at-risk and the necessary interest rates to apply in the case of consumer credit. One of our results is that the payment distribution will be heavy tailed.

3.1 Literature Review

Since the 1990s, credit risk modeling has received a lot of attention and interest. Throughout industry and academia, scholars have been studying and creating mathematical models to qualify, quantify and analyze credit risk. It has been and continues to be a major field of research and development. In particular, the consumer credit market has received a lot of study due to its size and significance in the world market. In the United States alone, consumer credit has grown massively in the past several decades reaching \$3.7 trillion at the end of 2016 [26]. Moreover, the consumer credit market was at the heart of the 2007-2009 financial crisis.

At the center of credit risk modeling are credit ratings, which play an important role in assessing an entity's ability to fulfill a financial commitment. In 1997, Jarrow, Lando and Turnbull [21] developed a Markov model that takes advantage of an entity's migration probabilities between different credit ratings. They introduced a discrete-time Markov chain model for the pricing and hedging of debt and bonds subject to default risk by incorporating credit rating migrations via a transition matrix. The main benefit of using a Markov model is that it allows for the specification of the distribution for the time of bankruptcy, which plays the main role in the determination of the pricing of a bond.

Since the introduction of Markov chains to credit risk modeling, a significant amount of research has taken place in assessing the main two assumptions made in the discrete-time Markov model presented by Jarrow et al. There has also been a lot of contributions in the literature for ways to take into account the lack of satisfaction of these assumptions. The first assumption is that a credit rating's migration is only dependent on the current rating; this is known as the memoryless property. The second assumption is the homogeneity assumption on the probabilities of migration. Jarrow et al. assume the Markov model to be homogenous, i.e. the probabilities of transition between the different ratings do not change over time. The lack of validity of these assumptions has been central to many papers in this field, and it will be central to this chapter.

Since Jarrow et al. published their paper, there have been a number of publications studying the assumptions in the Markov model. Nickell et al. (2000) use Moody's data from 1970 to 1997 to show that the transition matrix changes depending on the state of the economy [25]. Based on GDP growth, they place the economy into one of three states: peak, normal times and trough. Moreover, they find that the business cycle is the most important factor in causing the witnessed variation in the transition matrices. Independently, Bangia et al. (2002) use Standard & Poor's data from 1981 to 1998 to confirm these results by separating the economy into two states, expansion and contraction, and applying the model to credit portfolio stress testing [4]. They present a much more thorough numerical analysis on the statistics of the credit rating migration probabilities that were calculated from the data.

There have also been a number of papers that build on Nickell et al. and Bangia et al.'s analysis. Frydman and Schuermann [18] propose a model that is a mixture of two Markov chains that incorporates the speed of movement among the credit ratings. This is intended to accommodate for the lack of satisfaction of the memoryless property. Fitzpatrick and Marchev [16] present a double chain Markov model in which the transitions at any time are dependent on the unobservable state of the economy, which is modeled as a hidden Markov chain. Interestingly, some of their results are similar to the results presented by Bangia et al. A more recent 2016 paper in this regards is by D'Amico et al. [14], in which they model the migration as a semi-Markov process. This allows for the incorporation of the time spent in a state before migration.

The list of publications presented above have used similar mathematical frameworks to tackle the problem of assessing credit risk, but there have also been numerous papers by other scholars that have been studying credit risk with different numerical tools. In a recent 2016 publication by Butaru et al. [8], a more data driven machine learning approach is taken. They combine consumer historical, credit bureau and macroeconomic data to have a number of learning algorithms make predictions on the delinquency and default of credit card

accounts. They perform this study with the goal of assessing how different banks manage their risk and arrive at the conclusion that there is a range in how successful different banking institutions are at anticipating the delinquency and default of their customers' accounts.

While Butaru et al.'s models do a good job at predicting default, there is a fundamental difference between their approach and the Markov chain approaches mentioned above. In general, machine learning algorithms require a dense set of data in order to successfully come up with meaningful results. Given such a data set, the machine learning algorithm then "learns" the appropriate parameters that allows it to make correct predictions. While this makes such algorithms versatile, they often become complex "black boxes," where the causality that underlines the prediction is not clearly understood. On the other hand, the Markov model that we are presenting above stems from the fundamental structure of credit and its behavior. Moreover, adjustments to the original model are made to better capture and simulate the observed behavior. As a result, a Markov model can give a lot more insight into the casualty and behavior of the process that is being studied. That said, it would be hard to say that one modeling technique is better than the other, instead these two approaches complement each other and can play an important role as tools for a practitioner who, in the end, is the decision maker.

3.2 Discrete Time Markov Chain Approach

3.2.1 Markov Chain Model

Let $S = \{S_1, \ldots, S_K\}$ be the state space representing the different rating classes, with S_1 denoting the best credit rating and S_K denoting the absorbing default state, and let X_t be a random variable representing the rating of an entity that takes on values from S at time period t. Jarrow et al. assume that the process X_t is a Markov chain. Taking

$p_{t,ij} = P\left(X_{t+1} = S_j \mid X_t = S_i\right)$ gives rise to the $(K \times K)$ one period transition matrix

$$P_t = \begin{bmatrix} p_{t,11} & p_{t,12} & \cdots & p_{t,1K} \\ p_{t,21} & p_{t,22} & \cdots & p_{t,2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{t,K-1,1} & p_{t,K-1,2} & \cdots & p_{t,K-1,K} \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

where $p_{t,ij} \geq 0$ for all i, j and $\sum_{j=1}^K p_{t,ij} = 1$ for all i. We also assume $p_{t,ii} \neq 1$ for $1 \leq i < K$ and state S_K is reachable from all the other states. In Jarrow et al.'s model, this Markov chain is assumed to be homogeneous, and hence the transition matrix P is independent of t. In our economy switching model, we will assume that P_t will be taking on one of two matrix values: P_g for the time steps with economic expansion and P_b for the time steps with economic recession. However, we will begin by first exploring some properties of homogeneous Markov chains on a finite time interval of length T.

3.2.2 Expected Value of Visits to None-default States in a Finite Time Period

Proposition 3.2.1. *For a homogeneous Markov chain with transition matrix P written in the form*

$$P = \begin{bmatrix} Q & * \\ 0 & 1 \end{bmatrix},$$

let X^T be the random matrix whose i,j-th entry is the number of visits the chain makes to S_j given initial state S_i over the time frame $[0, T]$. Then the expected number of visits to S_j with initial state S_i is given by

$$\left[E\left[X^T\right]\right] = \sum_{\ell=0}^T Q^\ell.$$

Proof. Let $X_{ij}^{(\ell)}$ be a random variable which equals 1 if the chain, with initial state S_i, is in state S_j after ℓ steps, and equals 0 otherwise. Then

$$P\left(X_{ij}^{(\ell)} = 1\right) = q_{ij}^{(\ell)} \qquad \text{and} \qquad P\left(X_{ij}^{(\ell)} = 0\right) = 1 - q_{ij}^{(\ell)},$$

where $q_{ij}^{(\ell)}$ is the ij-th entry of Q^ℓ, with the notation $Q^0 = I$. Then

$$E\left[X_{ij}^{(\ell)}\right] = 1 \cdot q_{ij}^{(\ell)} + 0 \cdot \left(1 - q_{ij}^{(\ell)}\right) = q_{ij}^{(\ell)}$$

and hence the expected number of times the chain is in state S_j in the first T steps is

$$E\left[\sum_{\ell=0}^{T} X_{ij}^{(\ell)}\right] = \sum_{\ell=0}^{T} q_{ij}^{(\ell)}.$$

Let X_{ij}^T denote the number of visits to S_j given initial state S_i in time period $[0, T]$. Then

$$X_{ij}^T = \sum_{\ell=0}^{T} X_{ij}^{(\ell)}.$$

Note that $X_{ij}^T \neq X_{ij}^{(T)}$. Thus,

$$E\left[X_{ij}^T\right] = E\left[\sum_{\ell=0}^{T} X_{ij}^{(\ell)}\right] = \sum_{\ell=0}^{T} q_{ij}^{(\ell)}.$$

$\square$

Let Y_i^T denote the total time spent in default, given initial state S_i, up to time T. Then $Y_i^T = T - \sum_{j=1}^{K-1} X_{ij}^T$ and hence:

Corollary 3.2.1. $E\left[Y_i^T\right] = T - \sum_{j=1}^{K-1} E\left[X_{ij}^T\right]$

3.2.3 Distribution of Payments in a Finite Time Period

In our model, we will assume that for as long as an individual is not in default, he or she is making a payment. Let W_i^T denote the number of payments made up to time T, given initial state S_i. Note that $W_i^T = \sum_{j=1}^{K-1} X_{ij}^T$ and $W_i^T \in \{0, \ldots, T\}$. Given the transition matrix P, we may explicitly calculate the distribution of W_i^T.

Proposition 3.2.2. *Given the probability transition matrix P, written in the form*

$$P = \begin{bmatrix} Q & * \\ 0 & 1 \end{bmatrix},$$

we have,

$$\begin{bmatrix} P\left(W_1^T = k\right) \\ \vdots \\ P\left(W_{K-1}^T = k\right) \end{bmatrix} = Q^k \begin{bmatrix} p_{1K} \\ \vdots \\ p_{K-1,K} \end{bmatrix} \qquad \text{for } k = 0, \ldots T-1,$$

$$\begin{bmatrix} P\left(W_1^T = T\right) \\ \vdots \\ P\left(W_{K-1}^T = T\right) \end{bmatrix} = Q^T \mathbf{1}.$$

Proof. For $1 \le i \le K-1$ and $0 \le k \le T-1$, note that the event $W_i^T = k$ is equivalent to having the process with initial state S_i migrate between non-default states for a total of k steps, and then migrating into default. Moreover, we may express $W_i^T = k$ conditionally as

$$\begin{aligned} \{W_i^T = k\} &= \bigsqcup_{j=1}^{K-1} \{W_i^T = k \mid \text{process visits } S_j \text{ in the } k\text{-th step}\} \\ &= \bigsqcup_{j=1}^{K-1} \{\text{process visits } S_j \text{ in the } k\text{-th step then defaults}\}. \end{aligned}$$

Therefore,

$$P\left(W_i^T = k\right) \quad = \quad \sum_{j=1}^{K-1} p_{ij}^{(k)} p_{jK}.$$

As for the event $W_i^T = T$, this is equivalent to never defaulting. So,

$$\left\{W_i^T = T\right\} \quad = \quad \bigsqcup_{j=1}^{K-1} \left\{W_i^T = T \mid \text{process visits } S_j \text{ in the } T\text{-th step}\right\}$$

$$= \quad \bigsqcup_{j=1}^{K-1} \left\{\text{process vists } S_j \text{ in the final } T\text{-th step}\right\}$$

and hence

$$P\left(W_i^T = T\right) = \sum_{j=1}^{K-1} p_{ij}^{(T)}.$$

These two equalities give us what we want to prove when written in matrix notation. $\qquad\square$

3.3 Modeling Correlation

In this section, we will present our modeling of the correlation that takes place in the portfolio due to the shared economy over a short time frame $[0, T]$. The economy is assumed to start off healthy and to have a small $\varepsilon > 0$ probability of going into recession. Once the economy enters recession, it is assumed that it will stay in recession until the end of the timeframe of interest. We are effectively considering a homogenous Markov chain that may switch to another homogenous chain at each time step with a probability ε.

3.3.1 Distribution of the Number of Collected Payments

Let $W_{i,t}^T$ denote the number of payments of an individual made up to time T, given initial state S_i and assuming the economy changes states from expansion to recession at time t.

Then

$$0 \leq k \leq t-1: \quad \begin{bmatrix} P\left(W_{1,t}^T = k\right) \\ \vdots \\ P\left(W_{K-1,t}^T = k\right) \end{bmatrix} = Q_g^k \begin{bmatrix} (P_g)_{1K} \\ \vdots \\ (P_g)_{K-1,K} \end{bmatrix}$$

$$t \leq k \leq T-1: \quad \begin{bmatrix} P\left(W_{1,t}^T = k\right) \\ \vdots \\ P\left(W_{K-1,t}^T = k\right) \end{bmatrix} = Q_g^t Q_b^{k-t} \begin{bmatrix} (P_b)_{1K} \\ \vdots \\ (P_b)_{K-1,K} \end{bmatrix}$$

$$k = T: \quad \begin{bmatrix} P\left(W_{1,t}^T = k\right) \\ \vdots \\ P\left(W_{K-1,t}^T = k\right) \end{bmatrix} = Q_g^t Q_b^{k-t} \mathbf{1},$$

where Q_g and Q_b denote the sub-matrices of P_g and P_b, respectively. Let $f_{i,t}^T$ denote the distribution of an individual in an economy that changes states at time t. Then, $f_{i,t}^T(k) = P\left(W_{i,t}^T = k\right)$ for $0 \leq k \leq T$.

3.3.2 Convolutions

Let X be a random variable with mean μ and variance σ^2 with distribution f. Let $Y = \sum_{i=1}^n X_i$ where X_i are independent and identically distributed to X. Then the distribution of Y will be f^{n*}, the convolution of f with itself n items. Then, by the (Lindeberg-Levy) Central Limit Theorem, for large n

$$Z := \frac{Y - n\mu}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{(X_i - \mu)}{\sigma} \sim \mathcal{N}(0,1).$$

So, for large n,

$$Y = \sqrt{n}\sigma Z + n\mu \sim \mathcal{N}\left(n\mu, n\sigma^2\right).$$

Thus, for n large, we can expect the convolution f^{*n} to be close to the Gaussian with mean $n\mu$ and variance $n\sigma^2$.

3.3.3 Distribution of Payments in a Portfolio

Let f_t denote the distribution of the payments of one individual, given that the economy state changes states at time $t \in \{1, \ldots, T\}$. Then the distribution of one individual irrespective of when the economy changes states is given by

$$f = (1 - (T - 2)\varepsilon) f_T + \varepsilon \left(\sum_{t=1}^{T-1} f_t \right).$$

Suppose that we have a portfolio of size n. If we assume that we have a diversified portfolio, in which the behavior of one individual is assumed not to impact the behavior of another individual, then the distribution of the portfolio will simply be

$$\tilde{f}_1 = f^{n*}.$$

On the other hand, if we assume that the individuals are in the same economy, then this extra assumption will give us that the distribution of the portfolio will be

$$\tilde{f}_2 = (1 - (T - 2)\varepsilon) f_T^{n*} + \varepsilon \left(\sum_{t=1}^{T-1} f_t^{n*} \right),$$

where f_t^{n*} will be the distribution of the portfolio assuming the economy state changes at time t. Such a portfolio is an example of a non-diversified portfolio, in which the underlying behavior of the economy is impacting all the individuals in the same way.

For large n, $\tilde{f}_1$ will be close to a normal distribution with mean $n\mu$ and variance $n\sigma^2$. As for $\tilde{f}_2$, we expect it to be close to a sum of normal distributions and hence, for large enough n, it may have $T - 1$ peaks, where the one associated with f_T^{n*} will at times be the most dominant.

3.3.4 Calculating VaR

Let $f_i^{T,n}$ denote the distribution of a portfolio of size n with initial state S_i and a total of T periods. Given a confidence level $\alpha \in (0,1)$, define the VaR of the portfolio at the confidence level α as

$$\mathrm{VaR}_{i,\alpha}^{T,n} = \min\left\{0 \leq x \leq T : \sum_{j=x}^{T} f_i^{T,n}(j) \leq \alpha\right\}.$$

With α confidence, we expect that the total number of collected payments to be at least $\mathrm{VaR}_{i,\alpha}^{T,n}$.

3.3.5 Calculating Interest Rate

Let L denote the loan size in dollar. After calculating $v = \mathrm{VaR}_{i,\alpha}^{T,n}$, set the monthly payment as $x(v) = L/v$ and take $r(v) = \frac{T}{v} - 1$ to be the interest rate. With α confidence, the monthly payment of $x(v)$ will insure a profit and the associated interest rate will be $r(v)$. Indeed, to insure profit with α confidence, x must satisfy $xv = L$. Also, given payment size x and $R = 1 + r$, R must satisfy $LR = xnT$. So $r = R - 1 = \frac{nT}{v} - 1$.

3.3.6 Numerical Simulation

In our simulation, we will take the hypothetical transition matrices

$$P_g = \begin{bmatrix} .91879 & .07385 & .00718 & 0 & 0 & 0 & 0 & 0 \\ .01131 & .91264 & .07091 & .00308 & .00206 & 0 & 0 & 0 \\ .00102 & .02561 & .91189 & .05328 & .00615 & .00205 & 0 & 0 \\ 0 & .00206 & .05361 & .87938 & .05464 & .00825 & .00103 & .00103 \\ 0 & .00106 & .00425 & .04995 & .85122 & .07333 & .00425 & .01594 \\ 0 & .00109 & .00109 & .00543 & .05972 & .82193 & .02172 & .08903 \\ 0 & .00437 & .00437 & .00873 & .02511 & .05895 & .67795 & .22052 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$
P_b = \begin{bmatrix}
.2416 & .6463 & .0775 & .0269 & .0055 & .0014 & .0005 & .0003 \\
.0483 & .6279 & .2377 & .0619 & .0157 & .0053 & .0023 & .0008 \\
.0069 & .2132 & .4739 & .2148 & .0551 & .0204 & .0105 & .0052 \\
0 & .0495 & .2189 & .4654 & .1629 & .0570 & .0299 & .0165 \\
0 & .008 & .0478 & .2602 & .3811 & .1502 & .0883 & .0645 \\
0 & .0019 & .0083 & .0768 & .2735 & .3137 & .1856 & .1401 \\
0 & .0003 & .0033 & .0266 & .1440 & .2528 & .3257 & .2474 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix},
$$

and variables $\varepsilon = 0.1$ and $T = 12$. The distributions are plotted in Figure 3.3.1. As anticipated, for $n = 1$ we have the same distribution. For $n = 10$, we see a difference in the distributions of the portfolios initiating in states S_1, S_2, S_3, S_4 but the difference between the distributions of the portfolios initiating in S_5, S_6, S_7 is not significant. For $n = 100$, we start seeing a signifiant difference. We clearly see correlation in the non-diversified portfolio, while the diversified portfolio almost follows a normal distribution. In the non-diversified portfolio, we have flatter distributions with more probability mass to the right. Also, the non-diversified portfolio distributions have a heavier left tail. This remains to be the case for $n = 500$ and $n = 1000$. For $n = 500$ we begin to see several peaks, namely 8 peaks for portfolios with initial states S_1, S_2, S_3, S_4. For $n = 1000$, the peaks become more evident and more so for $n = 10000$.

In the case of a portfolio comprised of 100 individuals, we calculated the associated VaR and interest rates in Table 3.3.1. We see here that the calculated interest rates are within range of current credit market interest rates. Also, the extremely large interest rates for individuals in the lower states matches interest rates that are often charged by payday loans and other such loans.
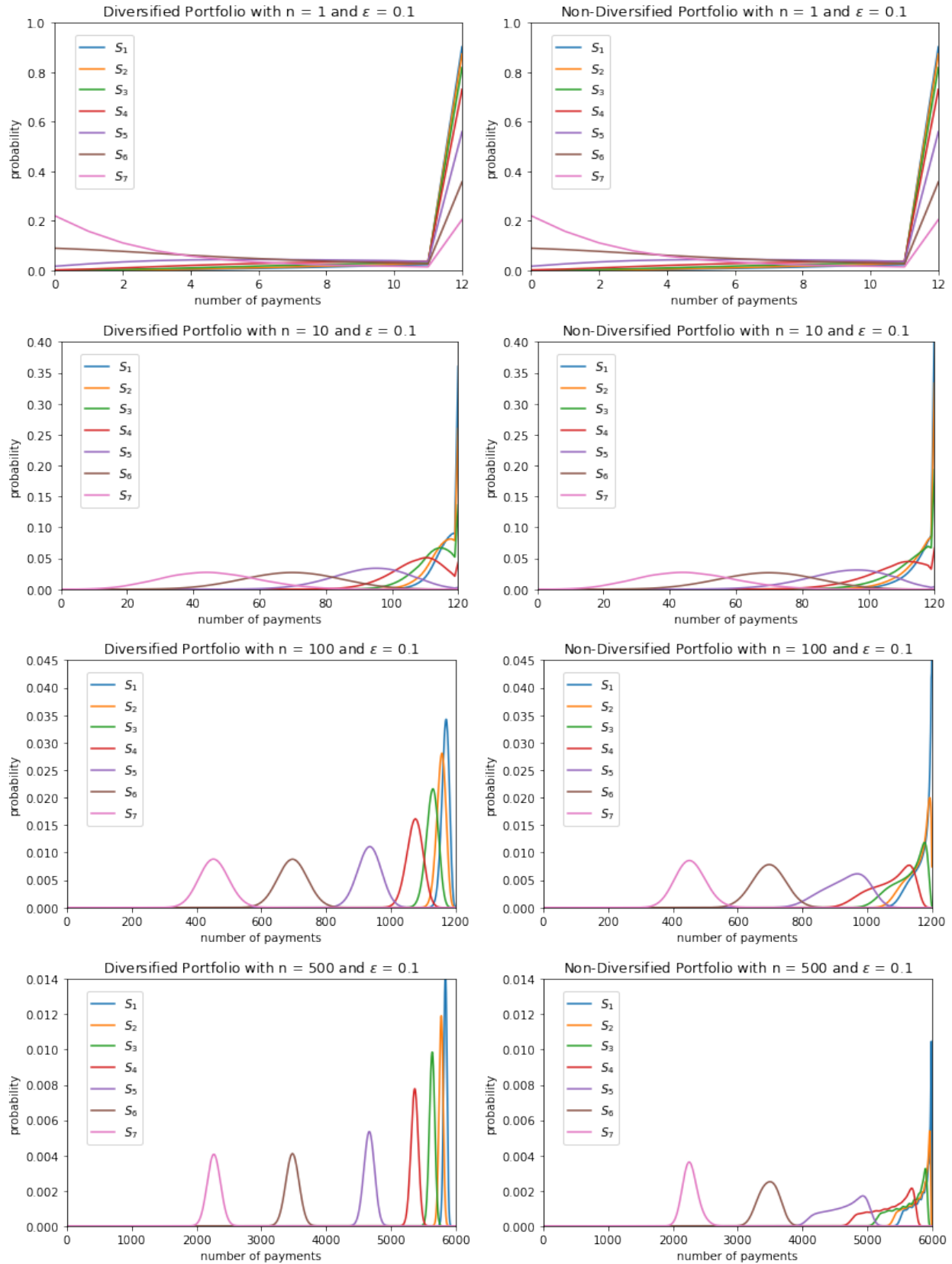
Figure 3.3.1: These plots represent the distributions of the number of payments in the diversified and non-diversified portfolios.

Initial State	Diversified Portfolio		Non-Diversified Portfolio	
	VaR	Interest Rate	VaR	Interest Rate
S_1	1147 (95.58%)	4.62%	1103 (91.92%)	8.79%
S_2	1130 (94.17%)	6.19%	1081 (90.08%)	11.01%
S_3	1097 (91.42%)	9.39%	1038 (86.5%)	15.61%
S_4	1036 (86.33%)	15.83%	968 (80.67%)	23.97%
S_5	879 (73.25%)	36.52%	825 (68.75%)	45.45%
S_6	626 (52.17%)	91.69%	617 (51.42%)	94.49%
S_7	384 (32.0%)	212.5%	382 (31.83%)	214.14%

Table 3.3.1: For $n = 100$, with 95% confidence, we calculated the VaR and interest rate in the diversified and non-diversified portfolios.

Chapter 4

Markov Chains with Random Transition Matrices

In a 1969 paper, Takahashi [33] presents a generalization of the ordinary Markov chain, which he calls a Markov chain with random transition matrices. He also presents some results in regards to the ergodicity of such Markov chains. In a more recent 2010 paper, Bruneau et al. [7] independently explore products of random transition matrices within a larger context and arrive at some similar, but different results. In this chapter, our goal is to shed light on Markov chains with random transition matrices by presenting an overview of this subject. We will present some of the material in these papers as well as explore some numerical examples to gain a better understanding of such chains and their applicability. One form of such chains that we will present is a perturbed Markov chain. Markov chains with random transition matrices are a very natural extension of ordinary Markov chains and can provide meaningful insight in applications where ordinary Markov chains are used.

4.1   Motivation and Some Examples

Let $\mathcal{P}$ denote a set of $s \times s$ transition matrices. A Markov chain with random transition matrices is a Markov process $\{X_n\}$ that is driven by a random matrix $P : \Omega \to \mathcal{P}$ over the

finite state space $S = \{1, \ldots, s\}$. Let

$$\alpha^{(0)} = \begin{bmatrix} \alpha_1^{(0)} & \cdots & \alpha_s^{(0)} \end{bmatrix}$$

denote the initial the distribution. Then the distribution of the Markov chain at time n is given by $\alpha^{(n)} = \alpha^{(0)} P^{(n)}$ where $P^{(n)} : \Omega^n \to \mathcal{P}$ is the transition matrix of such a chain over an n-step time period given by

$$P^{(n)}(\omega_1, \ldots, \omega_n) = \prod_{i=1}^{n} P(\omega_i).$$

In particular, we are assuming that $P(\omega_i)$ and $P(\omega_j)$ are independent for $i \neq j$. Thus,

$$\mathbb{E}\left[P^{(n)}\right] = \mathbb{E}\left[P\right]^n.$$

So, if $\mathbb{E}[P]^n$ converges for large n, the mean of $P^{(n)}$ will also converge. In order to better understand the behavior of such chains we will consider several examples.

Example 4.1.1. Let $S = \{1, 2, 3\}$ and let $\mathcal{P} = \{P_1, P_2\}$ given by

$$P_1 = \begin{bmatrix} .5 & .3 & .2 \\ .1 & .1 & .8 \\ .3 & .4 & .3 \end{bmatrix} \quad \text{and} \quad P_2 = \begin{bmatrix} .2 & .6 & .2 \\ .9 & .05 & .05 \\ .8 & .1 & .1 \end{bmatrix}$$

with $\Pr(P = P_i) = \frac{1}{2}$. Then P_1 and P_2 are irreducible and aperiodic with stationary distributions

$$\pi_1 \approx \begin{bmatrix} 0.3039 & 0.2843 & 0.4118 \end{bmatrix} \quad \text{and} \quad \pi_2 \approx \begin{bmatrix} 0.5215 & 0.3436 & 0.1350 \end{bmatrix}.$$
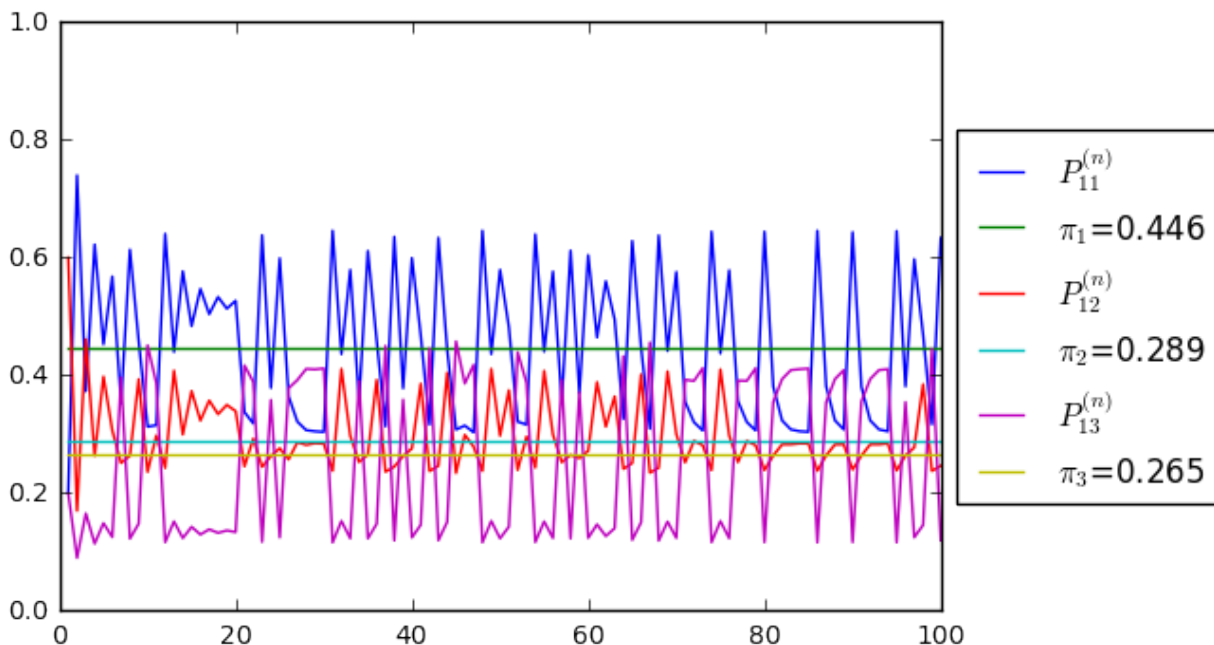
Figure 4.1.1: Plot of $P_{1j}^{(n)}$ in example 1 over $n = 1, \ldots, 100$ and the entries of the invariant distribution π of $\mathbb{E}[P]$.

Moreover,

$$\mathbb{E}[P] = \begin{bmatrix} .35 & .45 & .2 \\ .5 & 075 & .425 \\ .55 & .25 & .2 \end{bmatrix}$$

is also irreducible and aperiodic with the stationary distribution $\pi = \begin{bmatrix} 0.4463 & 0.2887 & 0.2650 \end{bmatrix}$.
In our simulation, we see that $P^{(n)}$ becomes rank 1, and its rows fluctuate around π. It is interesting to note that we have more fluctuation in the $P_{11}^{(n)}$ entry than in the $P_{12}^{(n)}$ entry, which presumably has to do with the interactions of P_1 and P_2.

In the next example we consider a particularly interesting Markov chain, where the chain is driven by a randomly perturbed matrix of the form $P + \varepsilon Q(\omega)$. Note that the randomness is only in Q. Such a chain will be called a perturbed Markov chain and we will discuss its construction in more detail below.
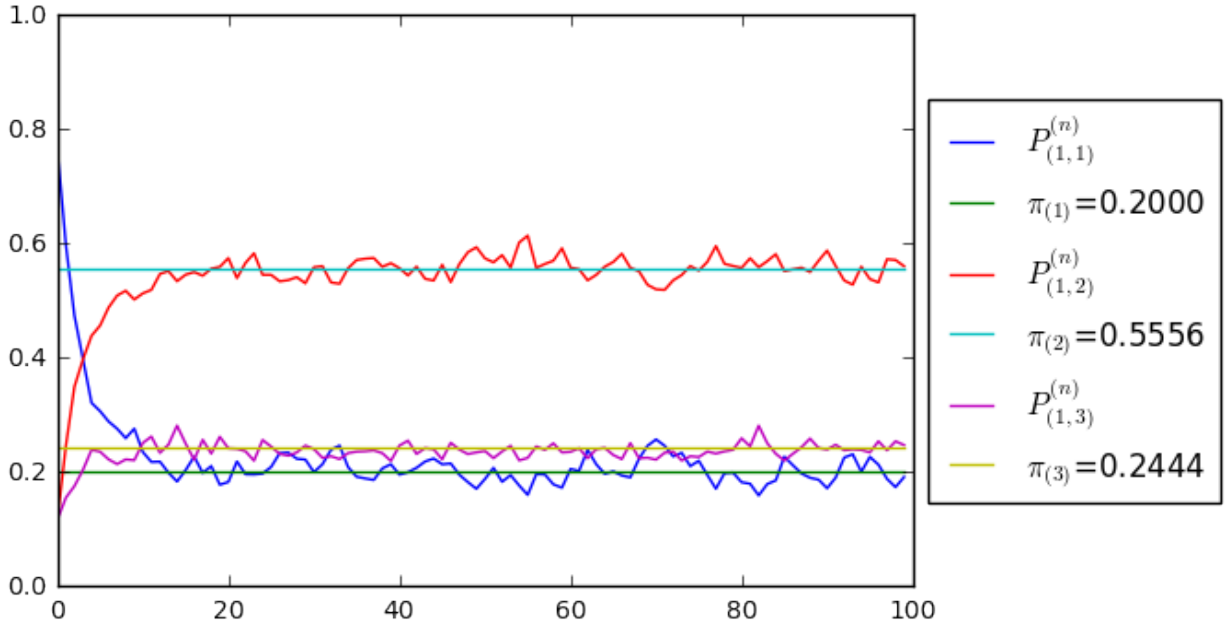
Figure 4.1.2: Plot of $P_{1j}^{(n)}$ in example 2 with $\varepsilon = 0.1$ over $n = 1, \ldots, 100$ and the entries of the invariant distribution π of P.

Example 4.1.2. Let $S = \{1, 2, 3\}$ and let $\mathcal{P} = \{P + \varepsilon Q\}$ where

$$
P = \begin{bmatrix} .8 & .1 & .1 \\ .05 & .7 & .25 \\ .05 & .6 & .35 \end{bmatrix},
$$

$\varepsilon \in (0, 1)$ is fixed and Q is random with mean 0 and rows summing to zero so that $P + \varepsilon Q$ remains a transition matrix. Note that $\mathbb{E}\left[P^{(n)}\right] = P^n$ because $\mathbb{E}[Q] = O$. As we saw in the above examples, $P^{(n)}$ will become rank 1 and will fluctuate around the invariant distribution $\pi = \begin{bmatrix} 1/5 & 5/9 & 11/45 \end{bmatrix} \approx \begin{bmatrix} 0.2 & 0.5556 & 0.2444 \end{bmatrix}$ of P for large n.

Another interesting example that can have many applications is a perturbed Markov chain with an absorbing state.
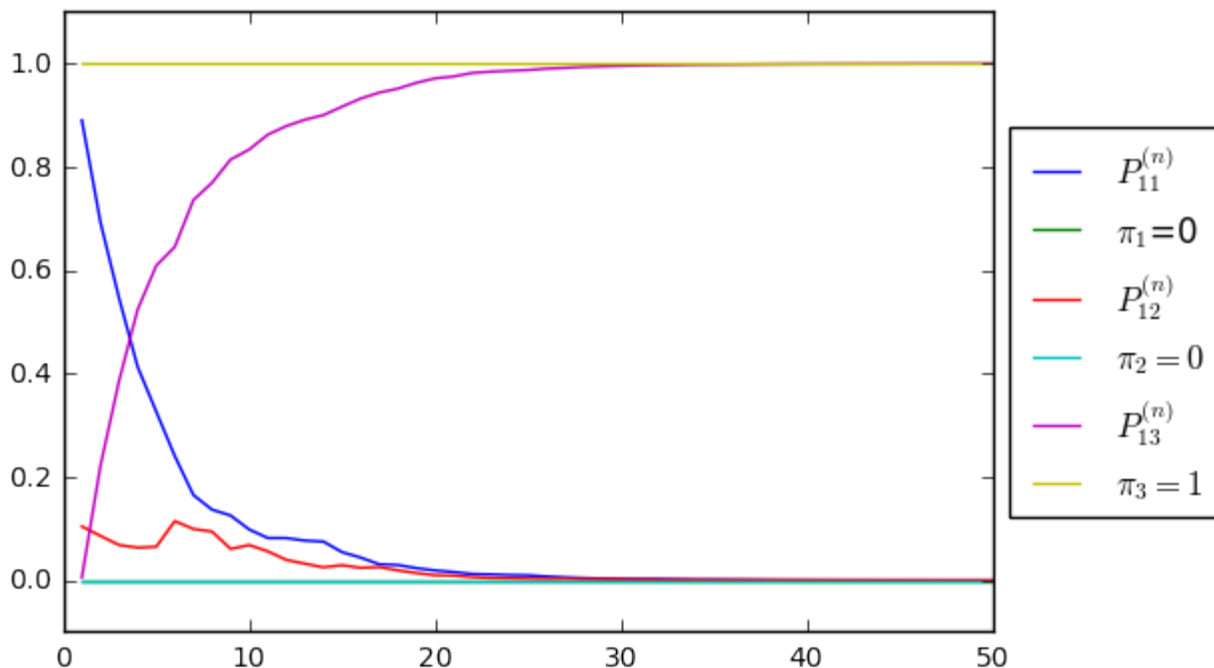
Figure 4.1.3: Plot of $P_{1j}^{(n)}$ in example 3 with $\varepsilon = 0.1$ over $n = 1, \ldots, 50$ and the entries of the invariant distribution π of P.

Example 4.1.3. Let $S = \{1, 2, 3\}$ and let $\mathcal{P} = \{P + \varepsilon Q\}$ where

$$
P = \begin{bmatrix} .8 & .1 & .1 \\ .05 & .7 & .25 \\ 0 & 0 & 1 \end{bmatrix},
$$

$\varepsilon \in (0, 1)$ is fixed and Q is random as above with the extra assumption that it does not alter the third row. Thus, all transition matrices in $\mathcal{P}$ will share an absorbing state. Then for any $\omega \in \Omega^n$,

$$
\lim_{n \to \infty} P^{(n)}(\omega) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.
$$

Indeed, we see this behavior in the simulation in Figure 4.1.3 where the entries fluctuate until the chain is fully absorbed.

37

4.2 Probabilistic Framework

Next we will present the probabilistic framework required to describe Markov chains with random transition matrices. Let $S = \{1, \ldots, s\}$ denote a finite state space, $\mathcal{A}$ be the set of all probability measures with respect to S and $\mathcal{P}$ be the set of all transition matrices with respect with to S. Then $\mathcal{A}$ and $\mathcal{P}$ can be viewed as s and s^2-dimensional Euclidean subspaces and we can denote them in the following fashion,

$$\mathcal{A} = \left\{ \begin{bmatrix} \alpha_1 & \cdots & \alpha_s \end{bmatrix} \in \mathbb{R}^s \mid \alpha_i \geq 0 \text{ and } \sum_i \alpha_i = 1 \text{ for all } i \right\}$$

and

$$\mathcal{P} = \left\{ [p_{ij}] \in \mathbb{R}^{s \times s} \mid p_{ij} \geq 0 \text{ and } \sum_k p_{ik} = 1 \text{ for all } i, j \right\}.$$

For each time step n, let $(\Omega^n, \mathscr{F}^n, \mathrm{Pr}^n)$ be a probability space for the $\mathcal{P}$-valued random matrix P^n and let

$$\left(\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathrm{Pr}} \right) = \prod_{n=1}^{\infty} (\Omega^n, \mathscr{F}^n, \mathrm{Pr}^n)$$

be the product probability space.

Definition 4.2.1. An S-valued *Markov chain $\{X_n\}$ with random transition matrices $\{P^n\}$* is a Markov process on $\overline{\Omega}$ such that

$$\mathrm{Pr}\left(X_n = j \mid X_0 = i \right) = \left(P^{(n)}\left(\overline{\omega} \right) \right)_{ij}$$

with $P^{(n)}\left(\overline{\omega} \right) = \prod_{i=1}^{n} P^i\left(\omega^i \right)$ and $\overline{\omega} = (\omega^1, \omega^2, \ldots) \in \overline{\Omega}$. In the particular case with $P^i = P$ for all i we will call the chain *stationary*.

Our main focus will be on stationary Markov chain with random transition matrices. It is worth noting that stationary Markov chains with random transition matrices are the analog of homogeneous Markov chains, while nonstationary Markov chain with random transition matrices are the analog of inhomogeneous Markov chains.

Let S_k be the set of k-tuples of states in S. Then the k-th moments of the random matrix $P(\omega)$ is the $s^k \times s^k$ matrix

$$\Sigma_k = \left[\mathbb{E}\left[p_{i_1 j_1} \cdots p_{i_k j_k} \right] \mid (i_1, \ldots, i_k), (j_1, \ldots, j_k) \in S_k \right].$$

We will adopt the following definitions for irreducibility, periodicity and ergodicity from Takahashi.

Definition 4.2.2. A stationary Markov chain with random transition matrices is *irreducible* if the ordinary Markov chain driven by Σ_1 is irreducible.

To define periodicity we will need to first present the following theorem, the proof of which can be found in Takahashi's paper as Theorem 4.

Theorem 4.2.1. *For a stationary and irreducible Markov chain with random transition matrices there exists a unique integer $r \geq 1$ such that the r sequences $\left\{ P^{(nr+m)} \right\}$ for $m \in \{0, 1, \ldots, r-1\}$ converge in distribution as n tends to ∞ and such that their limit distributions are different from each other.*

Definition 4.2.3. The *period* of a stationary and irreducible Markov chain with random transition matrices is the number r whose existence is assured in the above theorem. A stationary and irreducible Markov chain is *aperiodic* if $r = 1$.

Definition 4.2.4. A stationary and irreducible Markov chain with random transition matrices is *ergodic* if it is aperiodic and $P^{(n)}$ converges in distribution to a random matrix

$$Q = \begin{bmatrix} q_1 & \cdots & q_s \\ \vdots & \ddots & \vdots \\ q_1 & \cdots & q_s \end{bmatrix}.$$

Let $\hat{P}^{(n)}(\overline{\omega})$ denote the dual process of $P^{(n)}(\overline{\omega})$ in the sense that

$$\hat{P}^{(n)}(\overline{\omega}) = P(\omega_n) \cdots P(\omega_1).$$

Hence the distributions of $P^{(n)}(\overline{\varpi})$ and $\hat{P}^{(n)}(\overline{\varpi})$ coincide.

Lemma 4.2.1. *Let $\hat{M}_j^{(n)}(\overline{\varpi}) = \max_i \hat{p}_{ij}^{(n)}(\overline{\varpi})$. For each j, $\hat{M}_j^{(n)}(\overline{\varpi})$ converges to some $\hat{M}_j(\overline{\varpi})$.*

Define the random rank 1 matrix $\hat{M}$ by

$$
\hat{M}(\overline{\varpi}) = \begin{bmatrix} \hat{M}_1(\overline{\varpi}) & \cdots & \hat{M}_s(\overline{\varpi}) \\ \vdots & \ddots & \vdots \\ \hat{M}_1(\overline{\varpi}) & \cdots & \hat{M}_s(\overline{\varpi}) \end{bmatrix}.
$$

Theorem 4.2.2. *For a stationary and irreducible Markov chain with random transition matrices, the following four statements are equivalent:*

(a) The chain is ergodic.

(b) $P^{(n)}(\overline{\varpi})$ converges in distribution to $\hat{M}(\overline{\varpi})$.

(c) $P^{(n)}(\overline{\varpi})$ converges in probability to $\hat{M}(\overline{\varpi})$.

(d) $P^{(n)}(\overline{\varpi})$ converges with probability one to $\hat{M}(\overline{\varpi})$.

In the above results and others presented by Takahashi, the irreducibility of the Markov chain is an important and fundamental assumption. Bruneau et al study Markov chains with random transition matrices and infinite products of random matrices within the more general context of random dynamics processes and physics applications. Interestingly, they make a weaker assumption which produces similar results in regards to the ergodic behavior of such Markov chains. They show that the product $P^{(n)}$ is the sum of a fluctuating and a decaying process. The decaying process converges to zero almost surely and exponentially fast as $n \to \infty$. Instead of the irreducibility assumption, they assume that 1 is a simple eigenvalue of P.

Let $P(\omega)$ be as above, a $\mathcal{P}$-valued random matrix with simple eigenvalue 1, that is with multiplicity 1. Then for each ω, the Jordan form $J(\omega)$ exists along with an invertible $V(\omega)$,

such that

$$P(\omega) = V(\omega) J(\omega) U(\omega)$$

where $U(\omega) = V^{-1}(\omega)$ and $J(\omega)$ is of the form

$$J(\omega) = \begin{bmatrix} 1 & & & & \\ & B_2(\omega) & & & \\ & & \ddots & & \\ & & & B_p(\omega) \end{bmatrix}$$

where each matrix $B_i(\omega)$ is composed of the Jordan blocks of the eigenvalue $\lambda_i(\omega)$. Let

$$J_1 = \begin{bmatrix} 1 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 \end{bmatrix}$$

and

$$J_2(\omega) = \begin{bmatrix} 0 & & & & \\ & B_2(\omega) & & & \\ & & \ddots & & \\ & & & B_p(\omega) \end{bmatrix}.$$

Then we can decompose $P(\omega)$ as

$$P(\omega) = V(\omega) J_1 U(\omega) + V(\omega) J_2(\omega) U(\omega)$$

$$= \mathbf{1}\pi(\omega) + V(\omega) J_2(\omega) U(\omega)$$

where $\pi(\omega)$ is the invariant distribution of $P(\omega)$.

41

Proposition 4.2.1. $P^{(n)}(\omega_1, \ldots, \omega_n)$ *can be decomposed as follows:*

$$P^{(n)}(\omega_1, \ldots, \omega_n) = \mathbf{1}\pi(\omega_1) P^{(n-1)}(\omega_2, \ldots, \omega_n) + \prod_{i=1}^{n} V(\omega_i) J_2(\omega_i) U(\omega_i).$$

Proof. For $n = 2$,

$$P^{(2)}(\omega_1, \omega_2) = (\mathbf{1}\pi(\omega_1) + V(\omega_1) J_2(\omega_1) U(\omega_1)) P(\omega_2)$$

$$= \mathbf{1}\pi(\omega_1) P(\omega_2) + (P(\omega_1) - \mathbf{1}\pi(\omega_1)) \mathbf{1}\pi(\omega_2) + \prod_{i=1}^{2} V(\omega_i) J_2(\omega_i) U(\omega_i)$$

$$= \mathbf{1}\pi(\omega_1) P(\omega_2) + \prod_{i=1}^{2} V(\omega_i) J_2(\omega_i) U(\omega_i).$$

Given the above equation for some n,

$$P^{(n+1)}(\omega_1 \ldots, \omega_{n+1}) = \left[\mathbf{1}\pi(\omega_1) P^{(n-1)}(\omega_2, \ldots, \omega_n) + \prod_{i=1}^{n} V(\omega_i) J_2(\omega_i) U(\omega_i) \right] P(\omega_{n+1})$$

$$= \mathbf{1}\pi(\omega_1) P^{(n)}(\omega_2, \ldots, \omega_{n+1}) + \left(\prod_{i=1}^{n} V(\omega_i) J_2(\omega_i) U(\omega_i) \right) \mathbf{1}\pi(\omega_{n+1})$$

$$+ \prod_{i=1}^{n+1} V(\omega_i) J_2(\omega_i) U(\omega_i)$$

$$= \mathbf{1}\pi(\omega_1) P^{(n)}(\omega_2, \ldots, \omega_{n+1}) + \prod_{i=1}^{n+1} V(\omega_i) J_2(\omega_i) U(\omega_i).$$

By induction, the above decomposition holds for a general n. □

The following result, taken from Bruneau et al., shows that the second portion of the decomposition in Proposition 11 decays exponentially.

Theorem 4.2.3. *Let P be a $\mathcal{P}$-valued random matrix with positive probability of 1 being a simple eigenvalue. Then there exist a set $\Omega_1 \subset \overline{\Omega}$ and constants $C, \alpha > 0$ such that $\mathbb{P}(\Omega_1) = 1$ and such that for any $\overline{\omega} = (\omega_1, \omega_2, \ldots) \in \Omega_1$, there exists an $n_0(\overline{\omega})$ such that for*

any $n \geq n_0 (\overline{\omega})$,

$$\left\| \prod_{i=1}^{n} V(\omega_i) J_2(\omega_i) U(\omega_i) \right\| \leq C e^{-\alpha n}.$$

As a consequence of this theorem, $P^{(n)}(\overline{\omega})$ will become rank 1 for large n. In particular we have the following corollary.

Corollary 4.2.1. *Let P be a $\mathcal{P}$-valued random matrix with a constant invariant distribution π. Then*

$$\lim_{n \to \infty} P^{(n)}(\overline{\omega}) = \mathbf{1}\pi.$$

Proof. Indeed,

$$\mathbf{1}\pi P^{(n-1)}(\omega_2, \ldots, \omega_n) = \mathbf{1}\pi P(\omega_2) \cdots P(\omega_n)$$

$$= \mathbf{1}\pi$$

for all n since $\pi P(\omega_i) = \pi$ for all i. $\qquad\square$

When it comes to studying the fluctuation of the process, there are two perspectives we can take, both of which lead to the same result. On the one hand,

$$\mathbb{E}\left[P^{(n)}\right] = \mathbb{E}\left[P\right]^n$$

and hence the process will fluctuate about the invariant distribution of $\mathbb{E}[P]$ for large n. We can also study the Césaro mean. Bruneau et al. present the following result which proves the convergence in Césaro mean.

Theorem 4.2.4. *Let P be a $\mathcal{P}$-valued random matrix with simple eigenvalue 1 with positive probability. Then there exists a set $\Omega_2 \subset \overline{\Omega}$ such that $\mathbb{P}(\Omega_2) = 1$ and for any $\overline{\omega} \in \Omega_2$,*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P^{(n)}(\overline{\omega}) = \overline{\mathbb{E}[P]},$$

where

$$\overline{\mathbb{E}\left[P\right]} = \lim_{n \to \infty} \mathbb{E}\left[P\right]^n .$$

4.3   Perturbed Markov Chains

We are interested in exploring a specific type of Markov chain with random transition matrices, namely the perturbed Markov chain.

Definition 4.3.1. We will call a stationary Markov chain $\{X_n\}$ with random transition matrices $P + \varepsilon Q\left(\omega\right)$ a *perturbed Markov chain*, where P is a constant transition matrix, $\varepsilon \in (0,1)$ is constant and $Q\left(\omega\right)$ is a random matrix with rows summing to 0.

As we saw in examples 2 and 3, perturbed Markov chains have some interesting behaviors that we would like to explore and better understand. One specific property that we would like to better understand is the amount of fluctuation that takes place over time.

Let $\mathcal{P} = \{P + \varepsilon Q\left(\omega\right) \mid \omega \in \Omega\}$ be the set of perturbed transition matrices for some fixed $\varepsilon \in (0,1)$. Note that given a fixed P we will need to impose more assumptions on Q and ε to guarantee that $P + \varepsilon Q$ remains a transition matrix. We want to study inhomogeneous Markov chains driven by $\mathcal{P}$. The transition matrix of such a chain over an n-step time period is a random matrix $P^{(n)} : \overline{\Omega} \to \mathbb{R}^{s \times s}$ given by

$$P^{(n)}\left(\overline{\omega}\right) = \prod_{j=1}^{n} \left(P + \varepsilon Q\left(\omega_j\right)\right)$$

where $\overline{\Omega} = \prod_{j=1}^{\infty} \Omega$ and $\overline{\omega} = (\omega_1, \omega_2, \ldots) \in \overline{\Omega}$. To simplify the notation, we will let $Q_j := Q\left(\omega_j\right)$.

Let $L_\ell^{(n)}$ denote the order ε^ℓ term of $P^{(n)}$. Then

$$L_0^{(n)} = P^n,$$

$$L_1^{(n)} = \sum_{i=0}^{n-1} P^{n-1-i} Q_{n-i} P^i,$$

$$L_n^{(n)} = Q_1 \cdots Q_n,$$

$$L_\ell^{(n)} = \mathbb{O} \quad \text{for } \ell > n.$$

Moreover, we have the following recursion

$$L_\ell^{(n)} = L_\ell^{(n-1)} P + L_{\ell-1}^{(n-1)} Q_n$$

$$= L_\ell^{(\ell)} P^{n-\ell} + \sum_{i=0}^{n-\ell-1} L_{\ell-1}^{(n-1-i)} Q_{n-i} P^i$$

$$= Q_1 \cdots Q_\ell P^{n-\ell} + \sum_{i=0}^{n-\ell-1} L_{\ell-1}^{(n-1-i)} Q_{n-i} P^i.$$

4.3.1   Covariance of $P^{(n)}$ and $P^{(n+k)}$

In this section we will present an analytic expression for the limit of the covariance of $P^{(n)}$ and $P^{(n+k)}$ for a fixed k. In the proof, we will take advantage of the Jordan decomposition of P. If 1 is assumed to be a simple eigenvalue, then, as we have done earlier, we can decompose P as

$$P = \mathbf{1}\pi + V J_2 U.$$

Moreover, $P^i = \mathbf{1}\pi + V J_2^i U$ and $Q_j P^i = Q_j V J_2^i U$ for any i and j.

Proposition 4.3.1. *Let P be a transition matrix with simple eigenvalue 1 and let π be its invariant distribution. The limit of the entry-wise covariance of $P^{(n)}$ and $P^{(n+k)}$ is*

$$\lim_{n\to\infty} Cov\left(P_{ij}^{(n)}, P_{h\ell}^{(n+k)}\right) = \varepsilon^2 \left[\sum_{x=0}^{\infty} \sum_{a=1}^{s} \pi_a^2 \left((P^x - \mathbf{1}\pi)^T Cov(Q_{a\cdot}) \left(P^{x+k} - \mathbf{1}\pi\right) \right)_{j\ell} \right] + O\left(\varepsilon^3\right),$$

45

where $Cov(Q_{a\cdot})$ denotes the $s \times s$ covariance matrix of the a-th row of Q.

Proof. Recall that

$$\mathrm{Cov}\left(P_{ij}^{(n)}, P_{h\ell}^{(n+k)}\right) = \mathbb{E}\left[P_{ij}^{(n)} P_{h\ell}^{(n+k)}\right] - \mathbb{E}\left[P_{ij}^{(n)}\right] \mathbb{E}\left[P_{h\ell}^{(n+k)}\right].$$

By the independence assumptions,

$$\mathbb{E}\left[P_{ij}^{(n)}\right] \mathbb{E}\left[P_{h\ell}^{(n+k)}\right] = P_{ij}^n P_{h\ell}^{n+k}.$$

Note,

$$P_{ij}^{(n)} P_{h\ell}^{(n+k)} = \left(\sum_{x=0}^{n} \varepsilon^x L_{x,ij}^{(n)}\right) \left(\sum_{y=0}^{n+k} \varepsilon^y L_{y,h\ell}^{(n+k)}\right)$$

$$= L_{0,ij}^{(n)} L_{0,h\ell}^{(n+k)} + \varepsilon \left(L_{1,ij}^{(n)} L_{0,h\ell}^{(n+k)} + L_{0,ij}^{(n)} L_{1,h\ell}^{(n+k)}\right)$$

$$+ \varepsilon^2 \left(L_{2,ij}^{(n)} L_{0,h\ell}^{(n+k)} + L_{1,ij}^{(n)} L_{1,h\ell}^{(n+k)} + L_{0,ij}^{(n)} L_{2,h\ell}^{(n+k)}\right) + \cdots.$$

Calculating the expectation of the first few terms we have,

$$\mathbb{E}\left[L_{0,ij}^{(n)} L_{0,h\ell}^{(n+k)}\right] = P_{ij}^n P_{h\ell}^{n+k},$$

$$\mathbb{E}\left[L_{1,ij}^{(n)} L_{0,h\ell}^{(n+k)}\right] = \mathbb{E}\left[\left(\sum_{x=0}^{n-1} P^{n-1-x} Q_{n-x} P^x\right)_{ij} \left(P_{hl}^{n+k}\right)\right] = 0,$$

$$\mathbb{E}\left[L_{0,ij}^{(n)} L_{1,h\ell}^{(n+k)}\right] = 0,$$

$$\mathbb{E}\left[L_{2,ij}^{(n)} L_{0,h\ell}^{(n+k)}\right] = \mathbb{E}\left[\left(Q_1 Q_2 P^{n-2} + \sum_{x=0}^{n-3} \sum_{y=0}^{n-2-x} P^{n-2-x-y} Q_{n-1-x-y} P^y Q_{n-x} P^x\right)_{ij} \left(P_{h\ell}^{n+k}\right)\right] = 0$$

since $n - x > n - 1 - x - y$, and

$$\mathbb{E}\left[L_{0,ij}^{(n)} L_{2,h\ell}^{(n+k)}\right] = 0.$$

Now,

$$\mathbb{E}\left[L_{1,ij}^{(n)} L_{1,h\ell}^{(n+k)}\right] \overset{1}{=} \mathbb{E}\left[\left(\sum_{x=0}^{n-1} P^{n-1-x} Q_{n-x} P^x\right)_{ij} \left(\sum_{y=0}^{n+k-1} P^{n+k-1-y} Q_{n+k-y} P^y\right)_{h\ell}\right]$$

$$\overset{2}{=} \sum_{x,y} \mathbb{E}\left[\left(P^{n-1-x} Q_{n-x} P^x\right)_{ij} \left(P^{n+k-1-y} Q_{n+k-y} P^y\right)_{h\ell}\right]$$

$$\overset{3}{=} \sum_{x,y} \mathbb{E}\left[\sum_{a,b=1}^{s} \left(P_{ia}^{n-1-x} Q_{n-x,ab} P_{bj}^x\right) \sum_{c,d=1}^{s} \left(P_{hc}^{n+k-1-y} Q_{n+k-y,cd} P_{d\ell}^y\right)\right]$$

$$\overset{4}{=} \sum_{x,y} \sum_{a,b,c,d} \mathbb{E}\left[P_{ia}^{n-1-x} Q_{n-x,ab} P_{bj}^x P_{hc}^{n+k-1-y} Q_{n+k-y,cd} P_{d\ell}^y\right]$$

$$\overset{5}{=} \sum_{x=0}^{n-1} \sum_{a,b,d} \mathbb{E}\left[P_{ia}^{n-1-x} Q_{n-x,ab} P_{bj}^x P_{ha}^{n-1-x} Q_{n-x,ad} P_{d\ell}^{x+k}\right]$$

$$\overset{6}{=} \sum_{x=0}^{n-1} \sum_{a,b,d} \mathbb{E}\left[P_{ia}^{n-1-x} Q_{n-x,ab} \left(V J_2^x U\right)_{bj} P_{ha}^{n-1-x} Q_{n-x,ad} \left(V J_2^{x+k} U\right)_{d\ell}\right]$$

$$\overset{7}{=} \sum_{x=0}^{n-1} \sum_{a,b,d} P_{ia}^{n-1-x} P_{ha}^{n-1-x} \left(V J_2^x U\right)_{bj} \mathbb{E}\left[Q_{n-x,ab} Q_{n-x,ad}\right] \left(V J_2^{x+k} U\right)_{d\ell}$$

$$\overset{8}{=} \sum_{x=0}^{n-1} \sum_{a=1}^{s} P_{ia}^{n-1-x} P_{ha}^{n-1-x} \left(\sum_{b,d} \left(V J_2^x U\right)_{jb}^T \operatorname{Cov}\left(Q_{a\cdot}\right)_{bd} \left(V J_2^{x+k} U\right)_{d\ell}\right)$$

$$\overset{9}{=} \sum_{x=0}^{n-1} \sum_{a=1}^{s} P_{ia}^{n-1-x} P_{ha}^{n-1-x} \left(\left(V J_2^x U\right)^T \operatorname{Cov}\left(Q_{a\cdot}\right) \left(V J_2^{x+k} U\right)\right)_{j\ell}$$

$$\xrightarrow[n\to\infty]{10} \sum_{x=0}^{\infty} \sum_{a=1}^{s} \overline{P}_{ia} \overline{P}_{ha} \left(\left(V J_2^x U\right)^T \operatorname{Cov}\left(Q_{a\cdot}\right) \left(V J_2^{x+k} U\right)\right)_{j\ell}$$

$$\overset{11}{=} \sum_{x=0}^{\infty} \sum_{a=1}^{s} \pi_a^2 \left(\left(V J_2^x U\right)^T \operatorname{Cov}\left(Q_{a\cdot}\right) \left(V J_2^{x+k} U\right)\right)_{j\ell}$$

$$\overset{12}{=} \sum_{x=0}^{\infty} \sum_{a=1}^{s} \pi_a^2 \left(\left(P^x - \mathbf{1}\pi\right)^T \operatorname{Cov}\left(Q_{a\cdot}\right) \left(P^{x+k} - \mathbf{1}\pi\right)\right)_{j\ell}$$

where the fifth equality is due to the independence assumptions on the different Qs as well as

the rows of Q. Also we will have convergence in (10) since the magnitude of the summand decays geometrically. Indeed,

$$\left\| P_{ia}^{n-1-1} P_{ha}^{n-1-x} \left(V J_2^x U \right)^T \text{Cov} \left(Q_{a\cdot} \right) \left(V J_2^{x+k} U \right) \right\| = O \left(\lambda_2^{2x+k} \right)$$

where λ_2 is the second largest eigenvalue of P. $\square$

Corollary 4.3.1. *In the special case that the rows of Q are identically distributed, we have*

$$\lim_{n \to \infty} Cov \left(P_{ij}^{(n)}, P_{h\ell}^{(n+k)} \right) = \varepsilon^2 \left[\sum_{x=0}^{\infty} \pi \pi^T \left(P^x - \mathbf{1}\pi \right)^T Cov \left(Q_{1\cdot} \right) \left(P^{x+k} - \mathbf{1}\pi \right) \right]_{j\ell} + O \left(\varepsilon^3 \right).$$

Moreover, as $k \to \infty$, this term will decay exponentially.

Proof. Note that

$$\left\| \sum_{x=0}^{\infty} \pi \pi^T \left(P^x - \mathbf{1}\pi \right)^T \text{Cov} \left(Q_{1\cdot} \right) \left(P^{x+k} - \mathbf{1}\pi \right) \right\| = O \left(\lambda_2^k \right)$$

where λ_2 is the second largest eigenvalue of P, since $P^x - \mathbf{1}\pi = V J_2^x U$. $\square$

It is important to see here that λ_2 is governing the de-correlation in the process. Indeed, this turns out to be a powerful generalization from the classic Markov chain case which is causing a lot of the long term structure that we see with perturbed Markov chains.

Corollary 4.3.2. *In the special case that the rows of Q are identically distributed and $k = 0$, we have*

$$\lim_{n \to \infty} Cov \left(P_{ij}^{(n)}, P_{h\ell}^{(n)} \right) = \varepsilon^2 \left[\sum_{x=0}^{\infty} \pi \pi^T \left(P^x - \mathbf{1}\pi \right)^T Cov \left(Q_{1\cdot} \right) \left(P^x - \mathbf{1}\pi \right) \right]_{j\ell} + O \left(\varepsilon^3 \right).$$

This matrix will be symmetric and its diagonal gives the variance of $P_{ij}^{(n)}$, as $n \to \infty$.

We also have that the autocovariance of the process, i.e. the covariance of the process with itself over time, will decay to 0 in the limit.

Corollary 4.3.3. *Given the assumptions in Proposition 4.3.1,*

$$\lim_{k\to\infty} Cov\left(P_{ij}^{(n)}, P_{h\ell}^{(n+k)}\right) = \mathbb{O}$$

for any fixed n.

Proof. In the proof of Proposition 4.3.1, we see that

$$Cov\left(P_{ij}^{(n)}, P_{h\ell}^{(n+k)}\right) = \sum_{x=0}^{n-1}\sum_{a=1}^{s} P_{ia}^{n-1-x} P_{ha}^{n-1-x}\left((VJ_2^x U)^T \operatorname{Cov}(Q_{a\cdot})(VJ_2^{x+k}U)\right)_{j\ell}$$

$$\xrightarrow[k\to\infty]{} \sum_{x=0}^{n-1}\sum_{a=1}^{s} P_{ia}^{n-1-x} P_{ha}^{n-1-x}\left((VJ_2^x U)^T \operatorname{Cov}(Q_{a\cdot})(V\mathbb{O}U)\right)_{j\ell} = \mathbb{O}.$$

Moreover, this convergence will take place exponentially with rate corresponding to the second largest eigenvalue λ_2 of P since

$$\left\| P_{ia}^{n-1-1} P_{ha}^{n-1-x}(VJ_2^x U)^T \operatorname{Cov}(Q_{a\cdot})(VJ_2^{x+k}U)\right\| = O\left(\lambda_2^{2x+k}\right),$$

where k will dominate in the exponent. $\square$

4.3.2   Perturbed Markov Chain Simulations

In this section we will present a procedure for the numerical simulation of perturbed Markov chains. We will first present the Dirichlet distribution as well as some of its properties because we will be using it in the simulation of the random perturbation Q.

4.3.2.1   Dirichlet Distribution

The Dirichlet distribution can be interpreted as a distribution on probability distributions. The random vector $\theta = (\theta_1, \ldots, \theta_n)$ is a Dirichlet random variable if it has the probability density function

$$p_\alpha(\theta) = \frac{1}{B(\alpha)}\prod_{i=1}^{n}\theta_i^{\alpha_i-1}I_S(\theta)$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ is a constant vector with $\alpha_i > 0$, $S = \{x \in \mathbb{R}^n \mid \sum x_i = 1, x_i \geq 0\}$ is the probability simplex, and

$$B(\alpha) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)}{\Gamma(\alpha_0)}$$

is the multivariate Binomial function with $\alpha_0 = \sum_{i=1}^n \alpha_i$. Moreover, it holds that

$$\mathbb{E}[\theta_i] = \frac{\alpha_i}{\alpha_0},$$
$$\mathrm{Var}(\theta_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)},$$
$$\mathrm{Cov}(\theta_i, \theta_j) = \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0 + 1)} \quad \text{if } i \neq j.$$

We will use the notation $\theta \sim \mathrm{Dir}(\alpha)$ to denote a Dirichlet random vector.

There are two important advantages to using the Dirichlet distribution in our work. Firstly, it gives a reasonable approach to defining a random vector with nonnegative entries summing to 1. Via a translation and scaling of a Dirichlet random vector, we can define a random vector with bounded entries summing to 0. Indeed, the random vector

$$\gamma = c(\theta - \mathbb{E}[\theta])$$

has mean 0, entries bounded by c and summing up to 0. Secondly, the constant vector α allows for a range of behaviors. In particular, taking $\alpha = (1, \ldots, 1)$ gives the $(n-1)$-dimensional uniform distribution. The two quantities that determine the behavior of the distribution are α and α_0, sometimes called the base measure and concentration, respectively. The base measure α affects the mean, while the concentration α_0 affects the variance and covariance. A small α_0 gives rise to distributions with extreme behaviors, while letting $\alpha_0 \to \infty$ forces the variance and covariance to 0, and the samples of such a distribution will be clustered about a normalization of the base measure α.

4.3.2.2 Construction of Q

Let $\mathsf{q} = \{x = (x_1, \ldots, x_s) \in \mathbb{R}^s \mid \sum_i x_i = 0\}$ and let $q : \Omega \to \mathsf{q}$ be a random vector on a probability space $(\Omega, \mathcal{F}, \mathrm{Pr})$ with mean $\mathbb{E}[q_i] = 0$ and variance $\mathrm{Var}(q_i) = \sigma_i^2$. Then the covariance matrix of q is

$$\mathrm{Cov}(q) := \begin{bmatrix} \sigma_1^2 & \mathrm{Cov}(q_1, q_2) & \cdots & \mathrm{Cov}(q_1, q_s) \\ \mathrm{Cov}(q_2, q_1) & \sigma_2^2 & \cdots & \mathrm{Cov}(q_2, q_s) \\ \vdots & & & \vdots \\ \mathrm{Cov}(q_s, q_1) & \mathrm{Cov}(q_s, q_2) & \cdots & \sigma_s^2 \end{bmatrix}.$$

with the property that

$$0 = \sum_{i=1}^{s} \sigma_i^2 + 2 \sum_{i<j} \mathrm{Cov}(q_i, q_j)$$

since $\mathrm{Var}(\sum q_i) = 0$. Also, the rows (and columns) of $\mathrm{Cov}(q)$ sum to 0 since

$$\sum_{j=1}^{s} \mathrm{Cov}(q_i, q_j) = \sum_{j=1}^{s} \mathbb{E}[q_i q_j] = \mathbb{E}\left[q_i \sum_{j=1}^{s} q_j\right] = 0.$$

For the purposes of constructing an appropriate Q, we will want to impose two extra assumptions on the random vector q:

1. $|q_i| < \frac{1}{2}$ and

2. $\sum_{q_i > 0} q_i < \frac{1}{2}$.

Given s independent random vectors constructed as above, we can build the random matrix $Q : \Omega_1 \times \cdots \times \Omega_s \to \mathbb{R}^{s \times s}$ with rows $q^1, \ldots, q^s$.

One structure we can impose on these random vectors is the Dirichlet distribution. Let $\theta = (\theta_1, \ldots, \theta_s) \sim \mathrm{Dir}(\alpha)$ with $\alpha = (\alpha_1, \ldots, \alpha_s) \in \mathbb{R}^s$ and $\alpha_j > 0$. Taking $q = \frac{1}{2}(\theta - \mathbb{E}[\theta])$ as a row of Q gives us the three assumptions that we need:

1. $\displaystyle\sum_{j=1}^{s} q_j = \frac{1}{2} \sum_{j=1}^{s} \left(\theta_j - \frac{\alpha_j}{\alpha_0}\right) = \frac{1}{2}(1 - 1) = 0$

2. $|q_j| = \dfrac{1}{2}\left|\theta_j - \dfrac{\alpha_j}{\alpha_0}\right| < \dfrac{1}{2}$

3. $\displaystyle\sum_{j:q_j>0} q_j = \dfrac{1}{2}\sum_{j:\theta_j>\frac{\alpha_j}{\alpha_0}}\left(\theta_j - \dfrac{\alpha_j}{\alpha_0}\right) < \dfrac{1}{2}\sum_{j}\theta_j = \dfrac{1}{2}$

Let $\|\cdot\|$ denote the infinity norm on matrices defined by $\|A\| := \max_i \sum_j |A_{ij}|$. Then for any transition matrix P we have $\|P\| = 1$. Also for any matrix Q constructed as above we have $\|Q\| < 1$. Indeed, note that for any i,

$$\sum_j |Q_{ij}| = 2\sum_{j:Q_{ij}>0} Q_{ij} < 1.$$

4.3.2.3   Numerical Example

In the simulation of Example 4.1.2, we found the numerical variances to be $\mathrm{Var}(P_{11}^{(n)}) \approx 0.00052$, $\mathrm{Var}(P_{12}^{(n)}) \approx 0.00039$ and $\mathrm{Var}(P_{13}^{(n)}) \approx 0.00025$. According to the analytic result presented in Corollary 4.3.2, the variance is $\mathrm{Var}(P_{11}^{(n)}) = 0.00054$, $\mathrm{Var}(P_{12}^{(n)}) = 0.00041$ and $\mathrm{Var}(P_{13}^{(n)}) = 0.00024$.

Chapter 5

Absorbing Markov Chains with Random Transition Matrices

In this chapter, we will discuss absorbing Markov chains with random transition matrices. In the previous chapter we studied perturbed Markov chains within the context of stationary and irreducible Markov chains, as presented in the literature. A next natural step is to consider stationary absorbing Markov chains with random transition matrices in which the random transition matrices share absorbing states. We will begin our discussion by motivating this study, then laying the foundations and presenting some definitions and propositions. The primary theorem that we will present and prove is the entry-wise convergence of infinite products of random transition matrices with shared absorbing states. We will then present perturbed absorbing Markov chains.

5.1   Motivation

In the previous chapters, we motivated and discussed Markov chains with random transition matrices as present in the literature. We also developed on these concepts by presenting perturbed Markov chains, an analytic approximation of the fluctuation of such processes, as well as a methodology for simulating them. However, throughout that discussion our focus

was on matrices for which the eigenvalue 1 is simple. So we were predominantly looking at irreducible transition matrices. For such matrices we saw that the process has a decaying component as well as a fluctuating component. We were able to talk about chains with only one absorbing state as well and we saw that such chains decayed exponentially.

The next natural cases to consider are Markov chains with random absorbing transition matrices and perturbed Markov chains with more than one absorbing state. Transition matrices with more than one absorbing state don't satisfy the assumptions of the theorems previously discussed. To our knowledge, such processes have not been studied in the literature. Establishing them and proving results pertaining to their convergence may be very valuable due to the fact that absorbing Markov chains have numerous applications in the sciences. This extra level of randomness that we are introducing will allow for more complexity in such models. In order to gain some insight, consider the following example.

Example 5.1.1. Let $S = \{1, 2, 3, 4\}$ be the state space of a perturbed Markov chain and let $\mathcal{P} = \{P_0 + \varepsilon U(\omega)\}$ be the set of transition matrices driving a Markov process where $\varepsilon = 0.2$,

$$
P_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .3 & .3 \\ .1 & .4 & .4 & .1 \end{bmatrix}
$$

and U is simulated so that the non-absorbing rows are uniform and summing to 0. (This can be done as before with the Dirichlet distribution.) Then we see in Figure 5.1.1 that the product of random matrices taken from $\mathcal{P}$ will converge within a small region of the expected value. However, in each simulation the exact convergence differs. Moreover, we don't see the fluctuation that we saw in the previous chapter.
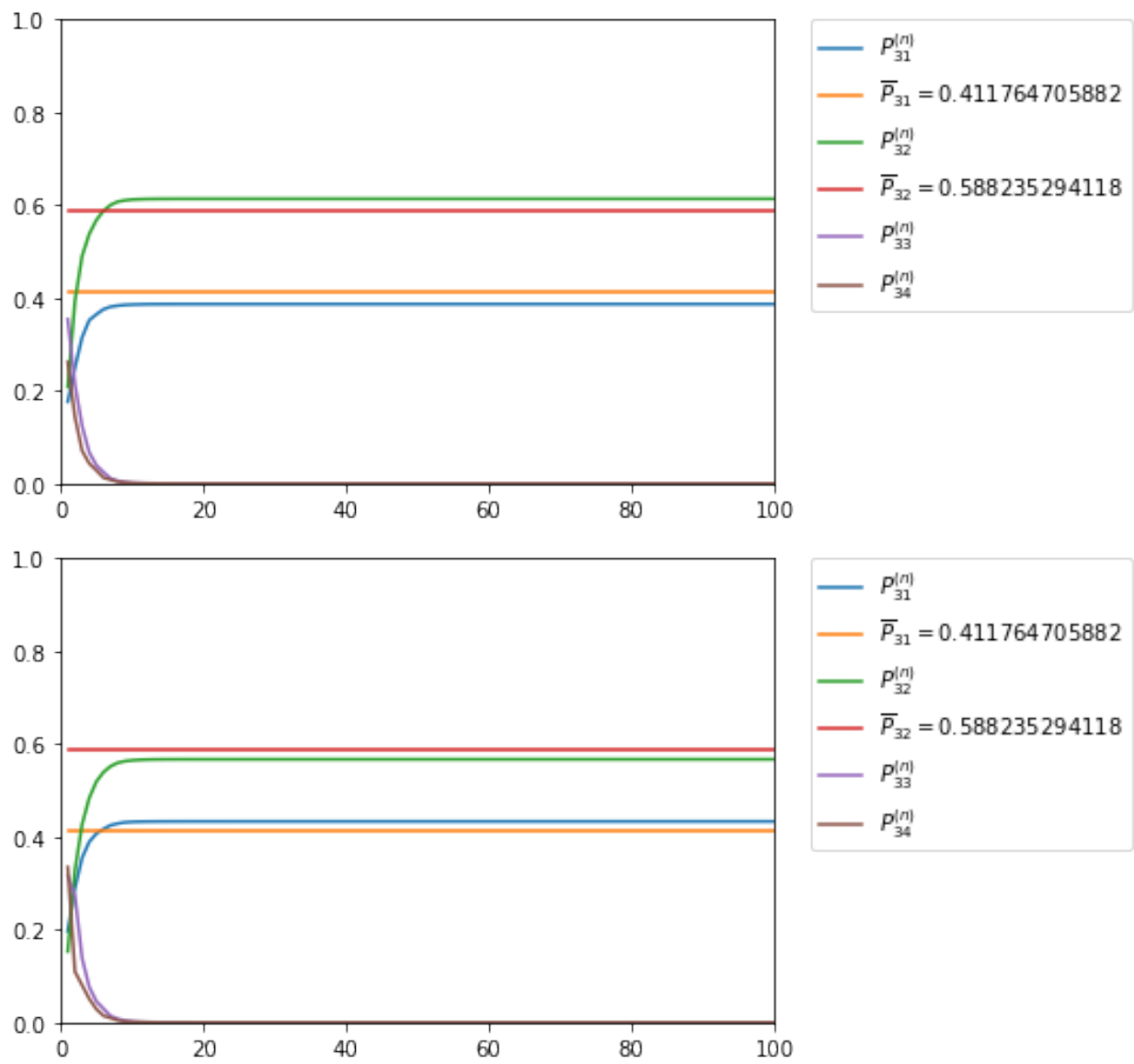
Figure 5.1.1: Here we have two simulations of the chain in example 5.1.1.

5.2 Probabilistic Framework

In this section, we will present the probabilistic framework and results that capture the asymptotic behavior of absorbing Markov chains with random transition matrices. Let $S = \{1, \ldots, s\}$ denote a finite state space and $\mathcal{P}^{(r)}$ be the set of all transition matrices with respect to S, where the first r states are absorbing states. Then

$$\mathcal{P}^{(r)} = \left\{ P = \begin{bmatrix} I_{r \times r} & O_{r \times (s-r)} \\ R_{(s-r) \times r} & Q_{(s-r) \times (s-r)} \end{bmatrix} \in \mathbb{R}^{s \times s} \,\middle|\, p_{ij} \geq 0 \text{ and } \sum_k r_{ik} + \sum_k q_{ik} = 1 \text{ for all } i, j \right\}.$$

Definition 5.2.1. An S-valued stationary r-absorbing Markov chain $\{X_n\}$ is a stationary Markov chain with random transition matrix $P(\omega) \in \mathcal{P}^{(r)}$.

All the primary results that were presented in the previous chapter relied on the irreducibility assumption. However, we will see in this section that some of the properties carry over. In the previous chapter, we saw that an infinite product of irreducible random transition matrices was composed of two parts, a geometrically decaying apart and a fluctuating part. These results relied on the assumption that 1 was a simple eigenvalue. In this chapter, we are studying absorbing transition matrices, for which the multiplicity of the eigenvalue 1 must match the number of absorbing states. However, we will see that infinite products of transition matrices with absorbing states will be composed of only a geometrically converging part. The goal of this section is to prove the following theorem.

Proposition 5.2.1. *Let $P(\omega)$ be a random transition matrix and suppose $P(\omega) \in \mathcal{P}^{(r)}$ with probability 1. Then there exists a set $\Omega_1 \subset \overline{\Omega}$ such that $\mathbb{P}(\Omega_1) = 1$ and such that for any $\overline{\omega} = (\omega_1, \omega_2, \ldots) \in \Omega_1$, the infinite product $\prod_{i=1}^{\infty} P(\omega_i)$ converges.*

5.2.1 Preliminary Deterministic Results

We will begin by first presenting several preliminary lemmas that hold true due to the deterministic properties of $P(\omega) \in \mathcal{P}^{(r)}$. For this reason, we will let P_i denote $P(\omega_i)$.

Lemma 5.2.1. *The spectral decomposition of P_i will be*

$$P_i = \begin{bmatrix} I & O \\ (I - Q_i)^{-1} R_i & O \end{bmatrix} + V_i \begin{bmatrix} O & O \\ O & \Lambda_i^{(2)} \end{bmatrix} V_i^{-1}$$

for some invertible matrix V_i and $\Lambda_i^{(2)}$ is a block diagonal square matrix comprised of the Jordan blocks associated with the eigenvalues that have magnitude less than 1.

Proof. Let $P_i \in \mathcal{P}^{(r)}$. First note that the algebraic multiplicity of the eigenvalue $\lambda_1 = 1$ is r since P_i has r absorbing states. Moreover, the geometric multiplicity of λ_1 must also be equal to r since

$$e_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T,$$

$$\vdots$$

$$e_r = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

are all eigenvectors with eigenvalue λ_1 and are linearly independent. Thus, the Jordan blocks associated with λ_1 are all 1 dimensional and hence P_i can be written in the form

$$P_i = V_i \begin{bmatrix} I & O \\ O & \Lambda_i^{(2)} \end{bmatrix} V_i^{-1}$$

$$= V_i \begin{bmatrix} I & O \\ O & O \end{bmatrix} V_i^{-1} + V_i \begin{bmatrix} O & O \\ O & \Lambda_i^{(2)} \end{bmatrix} V_i^{-1}$$

where $\Lambda_i^{(2)}$ is a block diagonal matrix comprised of the Jordan blocks of the eigenvalues $\lambda_2, \ldots, \lambda_p$ written in decreasing order with respect to their magnitude. Moreover,

$$P_i^n = V_i \begin{bmatrix} I & O \\ O & O \end{bmatrix} V_i^{-1} + V_i \begin{bmatrix} O & O \\ O & \left(\Lambda_i^{(2)}\right)^n \end{bmatrix} V_i^{-1}.$$

Since

$$P_i^\infty := \lim_{n \to \infty} P_i^n = \begin{bmatrix} I & O \\ (I - Q_i)^{-1} R_i & O \end{bmatrix}$$

and $\left(\Lambda_i^{(2)}\right)^\infty = O$, we must have

$$V_i \begin{bmatrix} I & O \\ O & O \end{bmatrix} V_i^{-1} = \begin{bmatrix} I & O \\ (I - Q_i)^{-1} R_i & O \end{bmatrix}.$$

Indeed, $\left(\Lambda_i^{(2)}\right)^\infty = O$ since $\left(\Lambda_i^{(2)}\right)^n$ will decay geometrically with respect to λ_2. □

In the above proof and throughout this chapter, we will let A^∞ denote the entry-wise limit $\lim_{n \to \infty} A^n$ for a square matrix A, assuming it exists. From here on, let

$$T_i = V_i \begin{bmatrix} O & O \\ O & \Lambda_i^{(2)} \end{bmatrix} V_i^{-1} = P_i - P_i^\infty.$$

In the probabilistic context, $T(\omega) = P(\omega) - P(\omega)^\infty$.

Lemma 5.2.2. *For all $i, j \in \mathbb{N}$,*

1. $P_i^\infty T_j = O$

2. $P_i^\infty P_j^\infty = P_i^\infty$

3. $P_i^\infty P_j = P_i^\infty$

Proof. The first two equations follow directly from simple matrix block-multiplication. The third equation follows from the first two since

$$P_i^\infty P_j = P_i^\infty \left(P_j^\infty + T_j\right) = P_i^\infty P_j^\infty + P_i^\infty T_j = P_i^\infty.$$

□

Lemma 5.2.3. *For $n \geq 2$,*

$$P_1 \cdots P_n = P_1^\infty + \sum_{i=1}^{n} T_1 \cdots T_{i-1} P_i^\infty + T_1 \cdots T_n.$$

Proof. We proceed with the proof inductively. For $n = 2$,

$$P_1 P_2 = (P_1^\infty + T_1) P_2$$
$$= P_1^\infty P_2 + T_1 P_2$$
$$= P_1^\infty + T_1 (P_2^\infty + T_2)$$
$$= P_1^\infty + T_1 P_2^\infty + T_1 T_2.$$

Now, suppose the above equality holds for some n. Then

$$P_1 \cdots P_{n+1} = \left(P_1^\infty + \sum_{i=1}^{n} T_1 \cdots T_{i-1} P_i^\infty + T_1 \cdots T_n \right) P_{n+1}$$
$$= P_1^\infty + \sum_{i=1}^{n} T_1 \cdots T_{i-1} P_i^\infty + T_1 \cdots T_n P_{n+1}$$
$$P_1^\infty + \sum_{i=1}^{n+1} T_1 \cdots T_{i-1} P_i^\infty + T_1 \cdots T_{n+1}.$$

$\square$

5.2.2   Probabilistic Results

Our next step is to prove the following proposition.

Proposition 5.2.2. *Let $P(\omega)$ be a random transition matrix and suppose $P(\omega) \in \mathcal{P}^{(r)}$ with probability 1. Then there exist constants $C, \alpha > 0$ such that for almost every $\overline{\omega} \in \overline{\Omega}$, there exists an $n_0(\overline{\omega})$ such that for any $n \geq n_0(\overline{\omega})$,*

$$\|T(\omega_1) \cdots T(\omega_n)\| \leq C e^{-\alpha n}.$$

Upon establishing Proposition 5.2.2, the proof of Proposition 5.2.1 will be as follows:

Proof. Take $\Omega_1 \subset \overline{\Omega}$ such that $\mathbb{P}(\Omega_1) = 1$ and for every $\overline{\omega} = (\omega_1, \omega_2, \ldots) \in \Omega_1$, the conclusion of Proposition 5.2.2 holds. Fix $\overline{\omega} \in \Omega_1$. To prove the convergence of the infinite product, it suffices to show that

$$\lim_{n \to \infty} \left\| \prod_{i=1}^{n} P(\omega_i) - \prod_{i=1}^{n-1} P(\omega_i) \right\| = 0.$$

Indeed, $\mathcal{P}^{(r)}$ can be viewed as a subset of the compact s^2 dimensional cube $[0, 1]^{s^2} \in \mathbb{R}^{s^2}$. By compactness, convergence of a sequence $\{x_n\}$ is equivalent to the convergence of $\{x_n - x_{n-1}\}$ to 0. Now, by Lemma 5.2.3

$$\left\| \prod_{i=1}^{n} P(\omega_i) - \prod_{i=1}^{n-1} P(\omega_i) \right\| = \| T(\omega_1) \cdots T(\omega_{n-1}) P(\omega_n)^{\infty}$$

$$+ T(\omega_1) \cdots T(\omega_n) - T(\omega_1) \cdots T(\omega_{n-1}) \|$$

$$\leq \| T(\omega_1) \cdots T(\omega_{n-1}) \| \, \| P(\omega_n)^{\infty} \|$$

$$+ \| T(\omega_1) \cdots T(\omega_n) \| + \| T(\omega_1) \cdots T(\omega_{n-1}) \|$$

$$\leq \| T(\omega_1) \cdots T(\omega_{n-1}) \| \, K$$

$$+ \| T(\omega_1) \cdots T(\omega_n) \| + \| T(\omega_1) \cdots T(\omega_{n-1}) \|,$$

which tends to 0 by Proposition 5.2.2. $\qquad \square$

In order to prove Proposition 5.2.2, we will first establish that for any ε we can approximate almost every $P(\omega)$ with finitely many matrices $P_1, \ldots, P_k \in \mathcal{P}^{(r)}$. Then the product $T(\omega_1) \cdots T(\omega_n)$ must decay geometrically bounded by the powers of the largest eigenvalue of the matrices P_i.

Proof. For a matrix $P \in \mathcal{P}^{(r)}$, let $\lambda_1 = 1, \lambda_2, \ldots, \lambda_p$ be its eigenvalues with $1 > |\lambda_2| > \cdots > |\lambda_p|$. For $n \in \mathbb{N}$, let

$$\mathcal{P}_n^{(r)} = \left\{ P \in \mathcal{P}^{(r)} \, \middle| \, |\lambda_2| \leq 1 - \frac{1}{n} \right\}.$$

Then the sequence of sets $\mathcal{P}_n^{(r)}$ is increasing and $\mathcal{P}^{(r)} = \bigcup_n \mathcal{P}_n^{(r)}$. Let $P(\omega) \in \mathcal{P}^{(r)}$ with probability 1. Then $P(\omega) \in \mathcal{P}_{n_0}^{(r)}$ for some n_0 with probability 1. Let $\varepsilon > 0$ and cover the closed set $\mathcal{P}_{n_0}^{(r)}$ with a finite number of balls with centers in $\mathcal{P}_{n_0}^{(r)}$ and radii ε. This is possible due to the compactness of $\mathcal{P}_{n_0}^{(r)}$ and the Heine–Borel property. Let $P_1, \ldots, P_k$ be the centers of these balls. Then for almost every $P(\omega)$ there exists an i such that $\|P(\omega) - P_i\| < \varepsilon$. Fix ω and i. Decompose $P_i = P_i^\infty + T_i$ and pick n_0 such that $\|P_i^{n_0} - P_i^\infty\| < \varepsilon/4$ and $\|P(\omega)^{n_0} - P(\omega)^\infty\| < \varepsilon/4$. By continuity of the function $A \mapsto A^{n_0}$, there exists $\delta > 0$ such that $\|P(\omega)^n - P_i^n\| < \varepsilon/4$ for $\|P(\omega) - P_i\| \le \delta$. Note that if $\varepsilon/4 < \delta$ we can replace δ by $\varepsilon/4$. Therefore,

$$
\begin{aligned}
\|T(\omega) - T_i\| &= \|(P(\omega) - P(\omega)^\infty) - (P_i - P_i^\infty)\| \\
&\le \|P(\omega) - P_i\| + \|P(\omega)^\infty - P_i^\infty\| \\
&\le \|P(\omega) - P_i\| + \|P(\omega)^\infty - P(\omega)^{n_0}\| + \|P(\omega)^{n_0} - P_i^{n_0}\| + \|P_i^{n_0} - P_i^\infty\| \\
&< \varepsilon.
\end{aligned}
$$

So, for any $\varepsilon > 0$, $T(\omega_j)$ can be decomposed as

$$
T(\omega_j) = T_j + \triangle(\omega_j)
$$

for some $T_j \in \{T_1, \ldots, T_k\}$ with $\|\triangle(\omega_j)\| < \varepsilon$, almost surely. Indeed, we are simply taking

$$
\triangle(\omega_j) = (P(\omega_j) - P(\omega_j)^\infty) - \left(P_j - P_j^\infty\right).
$$

Since we have only finitely many $P_j \in \mathcal{P}^{(r)}$, we can find positive constants C and γ such

that $\|T_1 \cdots T_n\| \leq Ce^{-\gamma n}$ for large n. By expanding the product, we have

$$\left\| \prod_{i=1}^{k} T(\omega_i) \right\| = \left\| \prod_{i=1}^{k} (T_i + \triangle(\omega_i)) \right\|$$

$$\leq \sum_{j=0}^{k} \binom{k}{j} Ce^{-\gamma(k-j)} \varepsilon^j$$

$$= Ce^{-\gamma k} \sum_{j=0}^{k} \binom{k}{j} (e^{\gamma}\varepsilon)^j = Ce^{-\gamma k} (1 + e^{\gamma}\varepsilon)^k .$$

For small enough ε, we can find $\alpha > 0$ so that $e^{-\gamma}(1 + e^{\gamma}\varepsilon) = e^{-\alpha}$. So we have that $\|T(\omega_1) \cdots T(\omega_k)\| \leq Ce^{-\alpha k}$ for almost every $(\omega_1, \ldots, \omega_k) \in \Omega^k$ and $k \in \mathbb{N}$. This establishes Proposition 5.2.2. $\qquad \square$

Before concluding this section, it is worth making a couple of remarks. Firstly, note that we need to have the strong assumption that almost every $P(\omega)$ has shared absorbing states. Indeed consider the example with

$$P(\omega) = \begin{cases} \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} & \text{with probability } \frac{1}{2} \\ \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix} & \text{with probability } \frac{1}{2} \end{cases} .$$

Then

$$P(\omega_1) \cdots P(\omega_n) = \begin{cases} \begin{bmatrix} 1 - q_{n,k} & q_{n,k} \\ 1 - p_{n,k} & p_{n,k} \end{bmatrix} & \text{with probability } \frac{1}{2^n} \\ \\ \begin{bmatrix} p_{n,k} & 1 - p_{n,k} \\ q_{n,k} & 1 - q_{n,k} \end{bmatrix} & \text{with probability } \frac{1}{2^n} \end{cases}$$

where $q_{n,k} = \frac{k}{2^{n-1}}$ and $p_{n,k} = \frac{2k-1}{2^n}$ for $k = 1, 2, \ldots, 2^{n-1}$. As a result, the infinite product will not converge and will fluctuate.

Secondly, we cannot allow $P(\omega)$ to be irreducible with positive probability because then we will have a fluctuating process. This is due to the results presented by Bruneau et al. However, we can relax the assumptions of Proposition 5.2.2 and allow for almost every $P(\omega)$ to lie in $\mathcal{P}^{(r+)} = \left\{ P \in \mathcal{P}^{(\ell)} \mid \ell \geq r \right\}$. Then the same argument on the closure of $\overline{\mathcal{P}^{(r+)}}$ will hold. That is, rather than finding a closed set $\mathcal{P}_{n_0}^{(r)}$, we can cover $\overline{\mathcal{P}^{(r+)}}$ with finitely many balls with radii ε. Taking the centers of these balls to be $P_1, \ldots, P_k$ will allow us to proceed in the same fashion.

5.3 Perturbed Absorbing Markov Chains

5.3.1 Definition

In this section we will discuss perturbed absorbing Markov chains. As in the previous sections, for $r, s \in \mathbb{N}$ with $r < s$, let

$$\mathcal{P}^{(r)} = \left\{ P = \begin{bmatrix} I_{r \times r} & O_{r \times (s-r)} \\ R_{(s-r) \times r} & Q_{(s-r) \times (s-r)} \end{bmatrix} \in \mathbb{R}^{s \times s} \,\middle|\, p_{ij} \geq 0 \text{ and } \sum_k r_{ik} + \sum_k q_{ik} = 1 \text{ for all } i, j \right\}.$$

Definition 5.3.1. A *perturbed absorbing Markov chain* is a perturbed Markov chain with random matrices $P(\omega) = P_0 + \varepsilon U(\omega)$, where $P_0 \in \mathcal{P}^{(r)}$, $\varepsilon \in (0, 1)$ and $U(\omega)$ is a random matrix with the first r rows taken to have 0 entries.

The random perturbations in such absorbing Markov chains do not impact the absorbing states. That is, if we write $U(\omega)$ as

$$U(\omega) = \begin{bmatrix} O & O \\ V(\omega) & W(\omega) \end{bmatrix},$$

then

$$P\left(\omega\right) = \begin{bmatrix} I & O \\ R\left(\omega\right) & Q\left(\omega\right) \end{bmatrix} = \begin{bmatrix} I & O \\ R_0 + \varepsilon V\left(\omega\right) & Q_0 + \varepsilon W\left(\omega\right) \end{bmatrix}.$$

As we saw before, the n-step time period transition matrix is a random matrix

$$P^{(n)} : \overline{\Omega} \to \mathbb{R}^{s \times s}$$

given by

$$P^{(n)}\left(\overline{\omega}\right) = \prod_{j=1}^{n} \left(P_0 + \varepsilon U\left(\omega_j\right)\right)$$

where $\overline{\omega} = \left(\omega_1, \omega_2, \ldots\right) \in \overline{\Omega}$. By proposition 5.2.1, $P^{(n)}\left(\overline{\omega}\right)$ converges in n almost surely.

5.3.2 Time Until Absorption

When studying classic absorbing chains, one of the important basic quantities that can be calculated is the expected time until absorption. If we have an absorbing chain driven by a transition matrix

$$P = \begin{bmatrix} I & O \\ R & Q \end{bmatrix},$$

then the expected time until absorption is given by $N\mathbf{1}$, where $N = I + Q + Q^2 + \cdots = (I - Q)^{-1}$ is the fundamental matrix of Q. In the context of inhomogeneous absorbing Markov chains driven by the sequence of matrices $P_1, P_2, \ldots$, the time until absorption is according to

$$N = I + \sum_{m=1}^{\infty} \prod_{k=1}^{m} Q_k = I + Q_1 + Q_1 Q_2 + \cdots .$$

In this subsection, our goal is to present an extension of the analysis within the context of perturbed absorbing Markov chains.

First note that for a sequence of random matrices $P\left(\omega_1\right), P\left(\omega_2\right), \ldots$, the matrix N is

random and dependent on $\omega_1, \omega_2, \ldots$ with

$$N(\omega_1, \omega_2, \ldots) = I + \sum_{m=1}^{\infty} \prod_{k=1}^{m} Q(\omega_k) = I + Q(\omega_1) + Q(\omega_1) Q(\omega_2) + \cdots.$$

Thus $\mathbb{E}[N] = \sum_{m=0}^{\infty} Q_0^m = (I - Q_0)^{-1}$ by independence. However, the variance is not as easily calculable and it would be beneficial to approximate the variance of the entries N_{ij}.

Proposition 5.3.1. *Let N be the random fundamental matrix of a perturbed absorbing Markov chain. Then*

$$Var(N_{ij}) = \varepsilon^2 \left(\sum_{m=1}^{\infty} \sum_{x=1}^{m} \left(Q_0^{x-1} \circ Q_0^{x-1} \right) C_x^{(m)} \right)_{ij} + O\left(\varepsilon^3 \right)$$

where the a-th row of $C_x^{(m)}$ is the diagonal of $Q_0^{m-x,T} Cov[W_{x,a\cdot}] Q_0^{m-x}$.

Proof. To simplify notation, fix a sequence $\omega_1, \omega_2, \ldots$ and let

$$Q_k = Q_0 + \varepsilon W_k$$

denote $Q(\omega_k) = Q_0 + \varepsilon W(\omega_k)$. Then

$$N = I + \sum_{m=1}^{\infty} \prod_{k=1}^{m} Q_k = \sum_{m=0}^{\infty} Q_0^m + \varepsilon \left(\sum_{m=1}^{\infty} \sum_{x=1}^{m} Q_0^{x-1} W_x Q_0^{m-x} \right) + O\left(\varepsilon^2 \right).$$

Let N_0 denote the fundamental matrix of Q_0. Then

$$N_{ij} N_{h\ell} = N_{0,ij} N_{0,h\ell} + \varepsilon \left[N_{0,ij} \left(\sum_{m=1}^{\infty} \sum_{x=1}^{m} Q_0^{x-1} W_x Q_0^{m-x} \right)_{h\ell} + \left(\sum_{m=1}^{\infty} \sum_{x=1}^{m} Q_0^{x-1} W_x Q_0^{m-x} \right)_{ij} N_{0,h\ell} \right]$$

$$+ \varepsilon^2 \left[N_{0,ij} (*) + \left(\sum_{m=1}^{\infty} \sum_{x=1}^{m} Q_0^{x-1} W_x Q_0^{m-x} \right)_{ij} \left(\sum_{n=1}^{\infty} \sum_{y=1}^{n} Q_0^{y-1} W_y Q_0^{n-y} \right)_{h\ell} + (*) N_{0,h\ell} \right]$$

$$+ O\left(\varepsilon^3 \right),$$

where the $(*)$ terms will not contain any powers of random matrices W_i and hence

$$\mathbb{E}\left[N_{0,ij}\left(*\right)\right] = \mathbb{E}\left[\left(*\right)N_{0,h\ell}\right] = 0.$$

So

$$Cov\left(N_{ij}, N_{h\ell}\right) = \varepsilon^2 \mathbb{E}\left[\left(\sum_{m=1}^{\infty}\sum_{x=1}^{m}Q_0^{x-1}W_xQ_0^{m-x}\right)_{ij}\left(\sum_{n=1}^{\infty}\sum_{y=1}^{n}Q_0^{y-1}W_yQ_0^{n-y}\right)_{h\ell}\right] + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{n=1}^{\infty}\sum_{y=1}^{n}\mathbb{E}\left[\left(Q_0^{x-1}W_xQ_0^{m-x}\right)_{ij}\left(Q_0^{y-1}W_yQ_0^{n-y}\right)_{h\ell}\right] + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{n=x}^{\infty}\mathbb{E}\left[\left(Q_0^{x-1}W_xQ_0^{m-x}\right)_{ij}\left(Q_0^{x-1}W_xQ_0^{n-x}\right)_{h\ell}\right] + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a,b,c,d}\mathbb{E}\left[Q_{0,ia}^{x-1}W_{x,ab}Q_{0,bj}^{m-x}Q_{0,hc}^{x-1}W_{x,cd}Q_{0,d\ell}^{k}\right] + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a,b,d}Q_{0,ia}^{x-1}Q_{0,ha}^{x-1}Q_{0,bj}^{m-x}\mathbb{E}\left[W_{x,ab}W_{x,ad}\right]Q_{0,d\ell}^{k} + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a,b,d}Q_{0,ia}^{x-1}Q_{0,ha}^{x-1}Q_{0,jb}^{m-x,T}\text{Cov}\left[W_{x,a\cdot}\right]_{bd}Q_{0,d\ell}^{k} + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a}Q_{0,ia}^{x-1}Q_{0,ha}^{x-1}\left(Q_0^{m-x,T}\text{Cov}\left[W_{x,a\cdot}\right]Q_0^{k}\right)_{j\ell} + O\left(\varepsilon^3\right).$$

In the case that we are interested in the variance, we have

$$\text{Var}\left(N_{ij}\right) = \varepsilon^2 \sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a}Q_{0,ia}^{x-1}Q_{0,ia}^{x-1}\left(Q_0^{m-x,T}\text{Cov}\left[W_{x,a\cdot}\right]Q_0^{k}\right)_{jj} + O\left(\varepsilon^3\right)$$

$$= \varepsilon^2 \left(\sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\left(Q_0^{x-1}\circ Q_0^{x-1}\right)C_x^{m,k}\right)_{ij} + O\left(\varepsilon^3\right)$$

where the a-th row of $C_x^{m,k}$ is the diagonal of $Q_0^{m-x,T}\text{Cov}\left[W_{x,a\cdot}\right]Q_0^{k}$. $\qquad\square$

While the above result gives some insight into the variance of the fundamental matrix, a

more applicable result in regards to the time until absorption is the following:

Corollary 5.3.1. *For a perturbed Markov chain with s states and r absorbing states, the expected time until absorption is given by the column vector* $t(\varpi) = N(\varpi)\mathbf{1}$. *Moreover,*

$$\mathbb{E}[t] = \mathbb{E}[N]\mathbf{1} = (I - Q_0)^{-1}\mathbf{1}$$

and

$$Var(t_i) = \varepsilon^2 \sum_{j=1}^{s-r}\sum_{\ell=1}^{s-r}\left[\sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a} Q_{0,ia}^{x-1}Q_{0,ia}^{x-1}\left(Q_0^{m-x,T}Cov[W_{x,a\cdot}]Q_0^k\right)_{j\ell}\right] + O\left((s-r)^2\varepsilon^3\right).$$

In particular, the standard deviation of t_i is of order $(s-r)\varepsilon$.

Proof. By the proof of the above proposition,

$$Var(t_i) = Var\left(\sum_{j=1}^{s-r}N_{ij}\right)$$

$$= \sum_{j=1}^{s-r}\sum_{\ell=1}^{s-r}Cov(N_{ij}, N_{i\ell})$$

$$= \varepsilon^2 \sum_{j=1}^{s-r}\sum_{\ell=1}^{s-r}\left[\sum_{k=0}^{\infty}\sum_{m=1}^{\infty}\sum_{x=1}^{m}\sum_{a} Q_{0,ia}^{x-1}Q_{0,ia}^{x-1}\left(Q_0^{m-x,T}Cov[W_{x,a\cdot}]Q_0^k\right)_{j\ell}\right] + O\left((s-r)^2\varepsilon^3\right).$$

$\square$

We confirm the calculation of the above proposition in the following example.

Example 5.3.1. Consider the perturbed absorbing Markov chain driven by random matrix $P = P_0 + \varepsilon Q(\omega)$ where

$$P_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.3 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 & 0.2 & 0.2 \\ 0.1 & 0.1 & 0.2 & 0.1 & 0.5 \end{bmatrix}$$

and $\varepsilon = 0.1$. Then the numerical approximation of the variance of N based on the above proposition is

$$\begin{bmatrix} 0.002625 & 0.001960 & 0.002648 \\ 0.002586 & 0.001931 & 0.002609 \\ 0.002696 & 0.002013 & 0.002720 \end{bmatrix},$$

while an approximation based on a large number of experiments is

$$\begin{bmatrix} 0.002596 & 0.001953 & 0.002683 \\ 0.002654 & 0.002010 & 0.002687 \\ 0.002716 & 0.002039 & 0.002776 \end{bmatrix}.$$

We see here that the two are equal up to 4 decimal places.

Chapter 6

Applications

In this chapter we will consider applications of absorbing Markov chains with random transition matrices.

6.1 Gambler's Ruin

6.1.1 Gambler's Ruin with Random Transition Matrices

In this section, we will generalize a classic application of absorbing Markov chains: the gambler's ruin problem. In the classic problem, a gambler starts with an initial fortune of i and then on each successive gamble either loses or wins \$1 with probabilities p and $1 - p$, respectively. He will keep playing until one of two events, he loses all his money or he makes \$N. If we let X_n denote the gambler's total fortune after the n-th gamble, then X_n can be viewed as an absorbing Markov chain on the state space $\mathcal{S} = \{0, 1, \ldots, N\}$ with transition

matrix

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

The long-term probabilities of winning or losing, given initial state i are P_{iN}^∞ and P_{i0}^∞. Moreover, it is a simple result that

$$P_{iN}^\infty = \begin{cases} \dfrac{1-\left(\frac{p}{1-p}\right)^i}{1-\left(\frac{p}{1-p}\right)^N} & \text{if } p \neq 0.5 \\[2ex] \dfrac{i}{N} & \text{if } p = 0.5 \end{cases}.$$

We would like to take this classic problem one step further by incorporating randomness to the transition matrix P. Instead of the constant p, we will consider a random variable u uniformly distributed on the unit interval. Then at each time step of the game, the gambler has different probabilities of migrating between the states. From the theory that we developed in the previous section, we can make several statements. First, for each gamble that the gambler plays, we are guaranteed longterm convergence in probabilities. This holds due to Proposition 5.2.1. Second, the longterm odds of the gambler will be largely determined by his odds in the initial rounds. That is, if the gambler's luck in the first few rounds is in his favor, the longterm probabilities will be in his favor above the expected probabilities. This is indeed true even if his luck turns against him after the first few rounds because the convergence of the process above or below its expectation is highly impacted by the initial draws of the matrix.

We ran some numerical simulations of this version of the gambler's ruin and we believe that the following conjecture holds.

Conjecture 6.1.1. *Consider an absorbing Markov chain driven by the random transition*

matrix

$$P = \begin{bmatrix} 1 & & & & \\ u & 0 & 1-u & & \\ & \ddots & \ddots & \ddots & \\ & & u & 0 & 1-u \\ & & & & 1 \end{bmatrix}$$

where u is uniformly distributed on the unit interval and n is the dimension of P. Then with $m = \left\lceil \frac{n}{2} \right\rceil - 1$ there exists parameters (α_i, β_i) for $i = 1, \ldots, m$ such that $P^{(n)}$ converges a.s. to the random matrix

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ Beta\,(\alpha_1, \beta_1) & 0 & \cdots & 0 & 1 - Beta\,(\alpha_1, \beta_1) \\ \vdots & 0 & \cdots & 0 & \vdots \\ Beta\,(\alpha_m, \beta_m) & 0 & \cdots & 0 & 1 - Beta\,(\alpha_m, \beta_m) \\ \vdots & 0 & \cdots & 0 & \vdots \\ 1 - Beta\,(\alpha_1, \beta_1) & 0 & \cdots & 0 & Beta\,(\alpha_1, \beta_1) \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix},$$

where $Beta\,(\alpha, \beta)$ denotes the Beta random variable with parameters α and β.

Figure 6.1.1 contains the histogram plots of the entries of $P^{(\infty)}$ with P a 5 dimensional matrix. We saw similar results for simulations of larger dimensions as well. In all our simulations, we were able to fit a beta distribution to our histogram.

In the 3-dimensional case, we can say the following:

Proposition 6.1.1. $P^{(n)}$ *converges to*

$$\begin{bmatrix} 1 & 0 & 0 \\ u & 0 & 1-u \\ 0 & 0 & 1 \end{bmatrix},$$
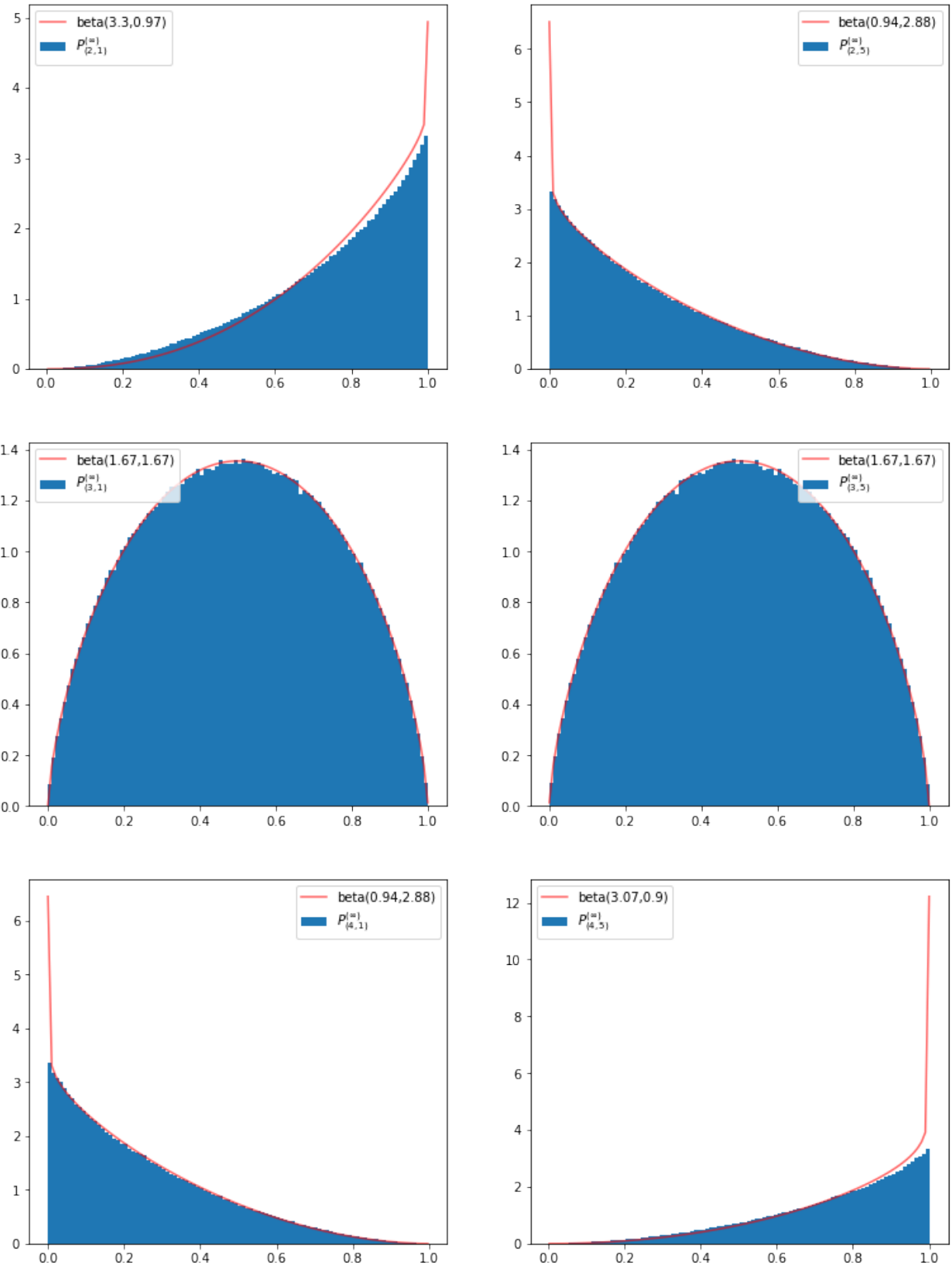
71

Figure 6.1.1: The histogram plots of the entries of $P^{(\infty)}$ with $N = 5$.

where u is the uniform random variable on the unit interval.

Note that the uniform distribution can also be viewed as a $Beta(1,1)$ distribution.

Proof. This trivially holds since

$$
\begin{bmatrix} 1 & 0 & 0 \\ u & 0 & 1-u \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ v & 0 & 1-v \\ 0 & 0 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ u & 0 & 1-u \\ 0 & 0 & 1 \end{bmatrix}
$$

for any u and v. $\qquad\square$

This result is interpretable within the context of our problem as only the first round will determine whether the gambler wins or loses and hence any more matrix draws will not change the probabilities.

In the 4-dimensional case, we believe the following holds:

Conjecture 6.1.2. $P^{(n)}$ *converges to*

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ t & 0 & 0 & 1-t \\ 1-t & 0 & 0 & t \\ 0 & 0 & 0 & 1 \end{bmatrix},
$$

where t is the triangular random variable with parameters $(0,1,1)$.

Note that the triangular random variable with parameters $(0,1,1)$ can also be viewed as a $Beta(2,1)$ random variable.

We may also look at the mean time of absorption. In the gambler's ruin, absorption takes place when all money is lost or the game is won in the two absorbing states 0 and N.
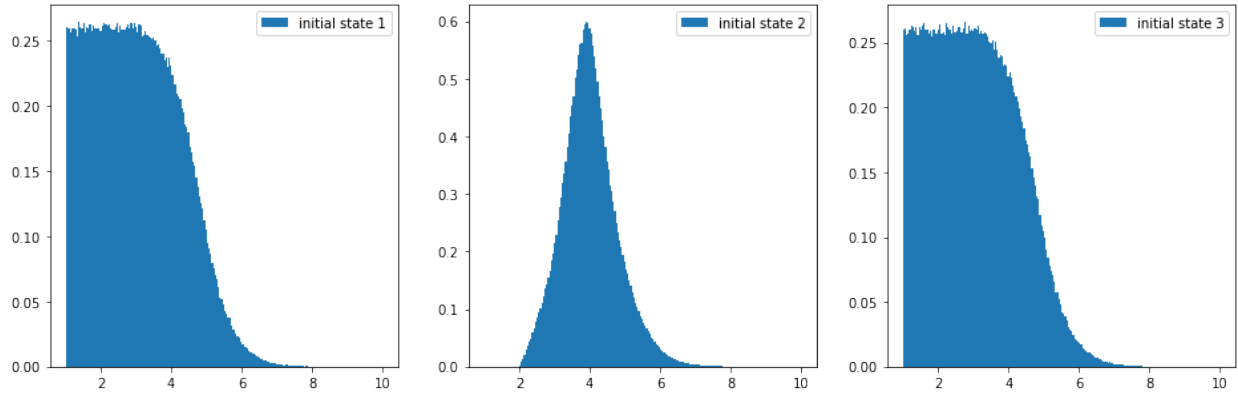
Figure 6.1.2: Histogram of mean time of absorption of gambler's ruin simulations with 4

In the classic example, this can be easily calculated. If we write

$$
P = \begin{bmatrix} 1 & 0 & 0 \\ A & T & B \\ 0 & 0 & 1 \end{bmatrix},
$$

then the i-th entry of the $N-1$ column vector $(I-T)^{-1}\mathbf{1}$ gives the mean time until absorption of the process that begins in state i.

In our context, the story develops complexity. Since the the probabilities are themselves random, the mean time until absorption is also random. In Figure 6.1.2, we plot the histogram of the time until absorption of simulations of the example started above with P a 5 dimensional matrix and $N = 4$. We see that the distribution of the mean time until absorption is the same for processes that begin in states 1 and 3 due to the symmetry in P. The probabilities of finishing in 1, 2 or 3 steps are the same, then the distribution develops a tail. The expectation of the mean time until absorption is 3 in these two cases. In the case that the process begins in state 2, the process needs a minimum of 2 steps and has an expectation of 4 steps. The distribution again develops a tail for larger values. We see similar histograms in simulations with other values of N. In Figure 6.1.3 we plot the histogram of the $N = 6$ case.
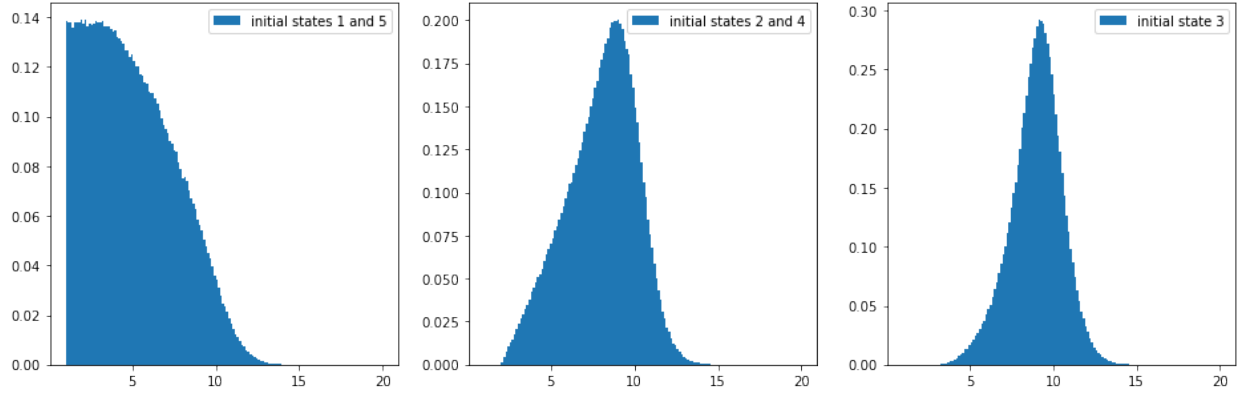
74

Figure 6.1.3: Histogram of mean time of absorption of gambler's ruin simulations with $N = 6$

6.1.2 Perturbed Gambler's Ruin

In this section, we will consider a generalization of the gambler's ruin problem within the context of perturbed Markov chains. Take

$$
P_0 = \begin{bmatrix}
1 & & & & & \\
p & 0 & 1-p & & & \\
& \ddots & \ddots & \ddots & & \\
& & p & 0 & 1-p \\
& & & & 1
\end{bmatrix}
$$

and

$$
Q = \begin{bmatrix}
0 & & & & & \\
u_1 & 0 & -u_1 & & & \\
& \ddots & \ddots & \ddots & & \\
& & u_{N-1} & 0 & -u_{N-1} \\
& & & & 0
\end{bmatrix}
$$

where u_i is uniformly distributed on the interval $(-1, 1)$. Let $P = P_0 + \varepsilon Q$ with $0 \leq \varepsilon \leq \min\{p, 1-p\}$. We would like to study the behavior of $P^{(\infty)}(\overline{\omega})$ and the mean time until absorption.

75

Note that the current formulation of the gambler's ruin is equivalent to saying that P is of the form

$$P = \begin{bmatrix} 1 & & & & \\ w_1 & 0 & 1-w_1 & & \\ & \ddots & \ddots & \ddots & \\ & & w_{N-1} & 0 & 1-w_{N-1} \\ & & & & 1 \end{bmatrix}$$

with w_i uniformly distributed on $(p-\varepsilon, p+\varepsilon)$. However, decomposing P as above will allow us to say more about the process and apply the results that were developed in the previous section.

Consider a perturbed gambler's ruin with $p = \frac{1}{2}$. In Figures 6.1.4 and 6.1.5 we plot the histograms of the simulations that we ran for a perturbed gambler's ruin with $\varepsilon = 0.2$ and $\varepsilon = 0.5$. In these two figures, we see the distribution of the entries of $P^{(\infty)}$ and how the increase in ε is leading to to a wider distribution. In Figures 6.1.6 and 6.1.7 we plot the histograms of the mean time of absorption for the two cases. The mean values of the mean time to absorption are the same for all ε. The standard deviation of the mean time to absorption for the three distributions in Figure 6.1.6 are 1.02, 0.86 and 0.72. Similarly, the standard deviation of the mean time to absorption for the three distributions in Figure 6.1.7 are 2.72, 2.47 and 2.21. This confirms the conclusion of Corollary 5.4.1 that the standard deviation will be of order $\varepsilon(t-s)$, which is 1 and 2.5 for $\varepsilon = 0.2$ and 0.5, respectively.

We also ran simulations and plotted the histograms of the time to absorption for the classic gambler's ruin and the perturbed gambler's ruin with $\varepsilon = 0.2$ and 0.5 in Figure 6.1.8 from left to right. The blue plots are the distributions of the simulations, with their mean in green. The red plot is the distribution of the classic gambler's ruin case. We see see in these plots that the mean distribution of the time until absorption in the random case and non-random case coincide.
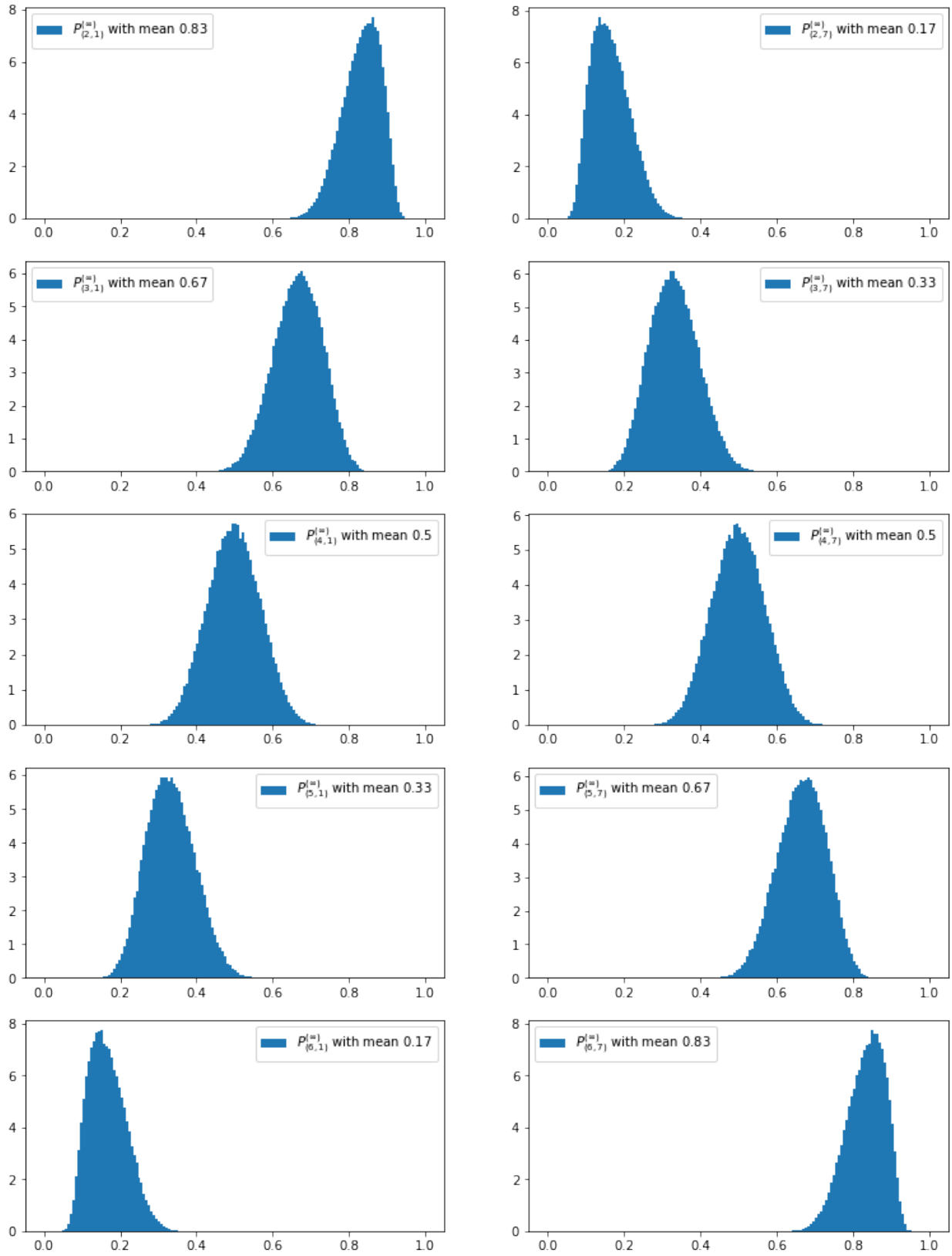
Figure 6.1.4: Histogram of $P^{(\infty)}$ of perturbed gambler's ruin simulations with $N = 6$ and $\varepsilon = 0.2$

Figure 6.1.5: Histogram of $P^{(\infty)}$ of perturbed gambler's ruin simulations with $N = 6$ and $\varepsilon = 0.5$

Figure 6.1.6: Histogram of mean time of absorption of perturbed gambler's ruin simulations with $N = 6$ and $\varepsilon = 0.2$
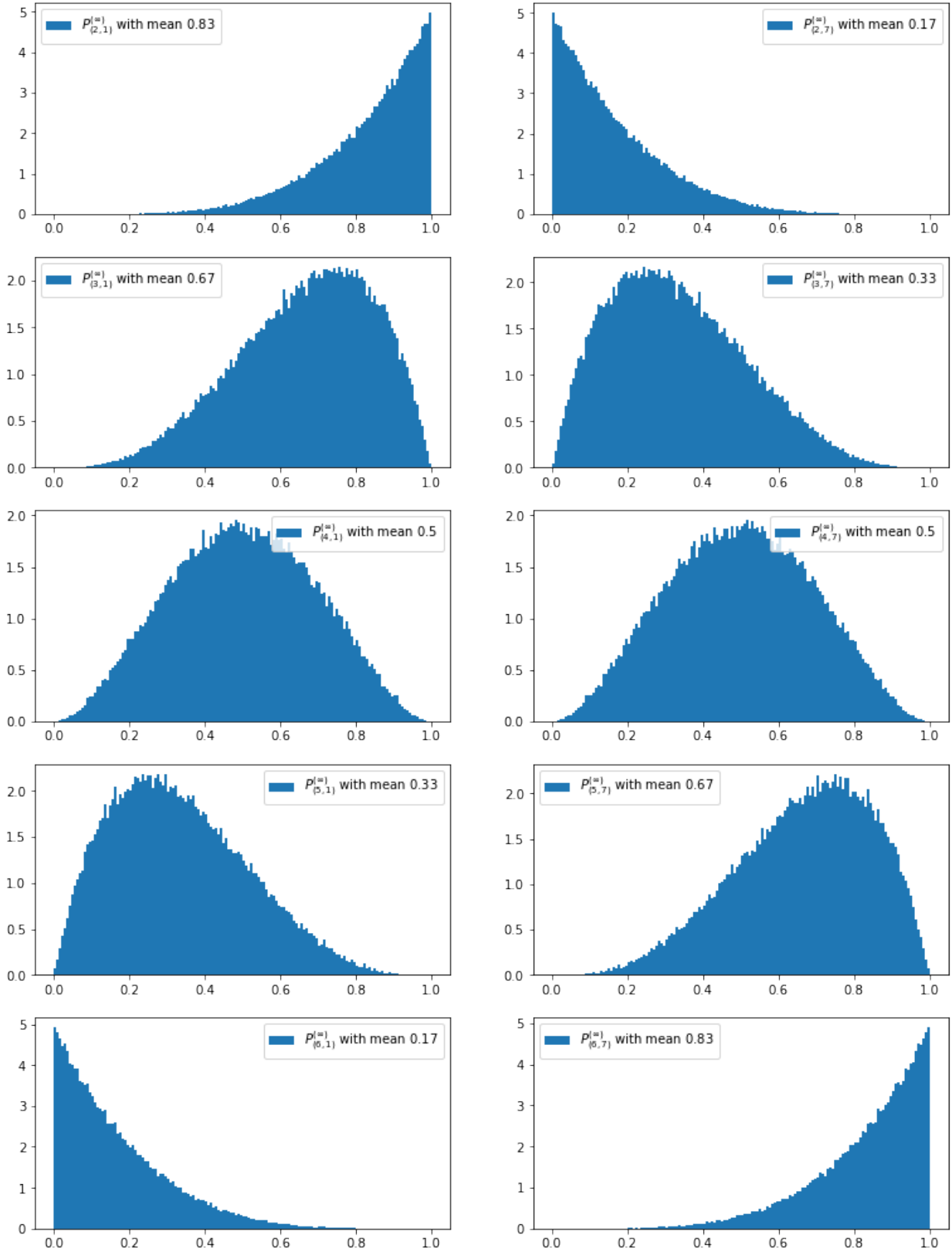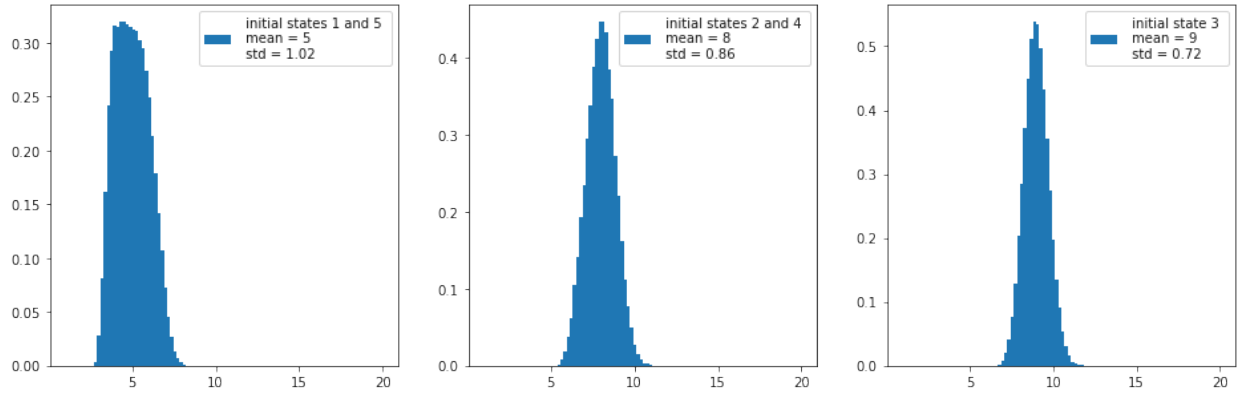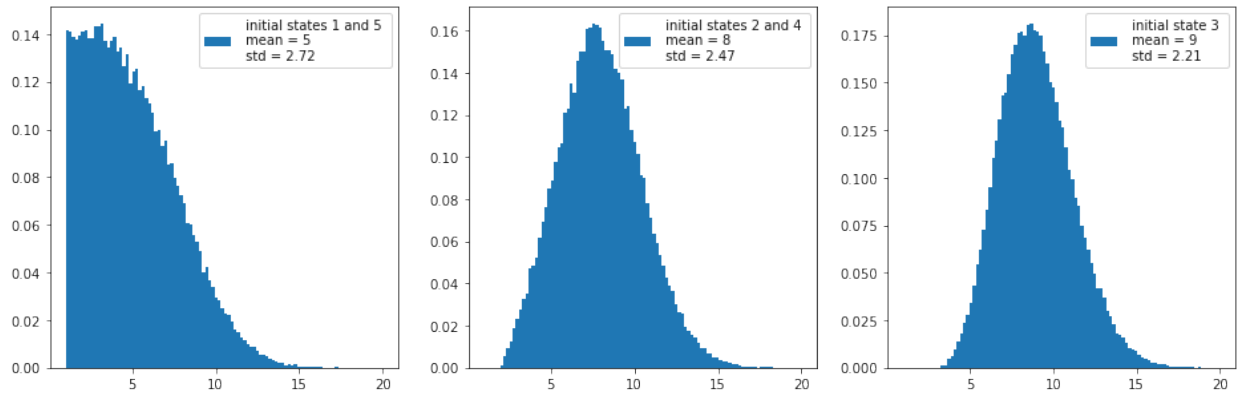

Figure 6.1.7: Histogram of mean time of absorption of perturbed gambler's ruin simulations with $N = 6$ and $\varepsilon = 0.5$
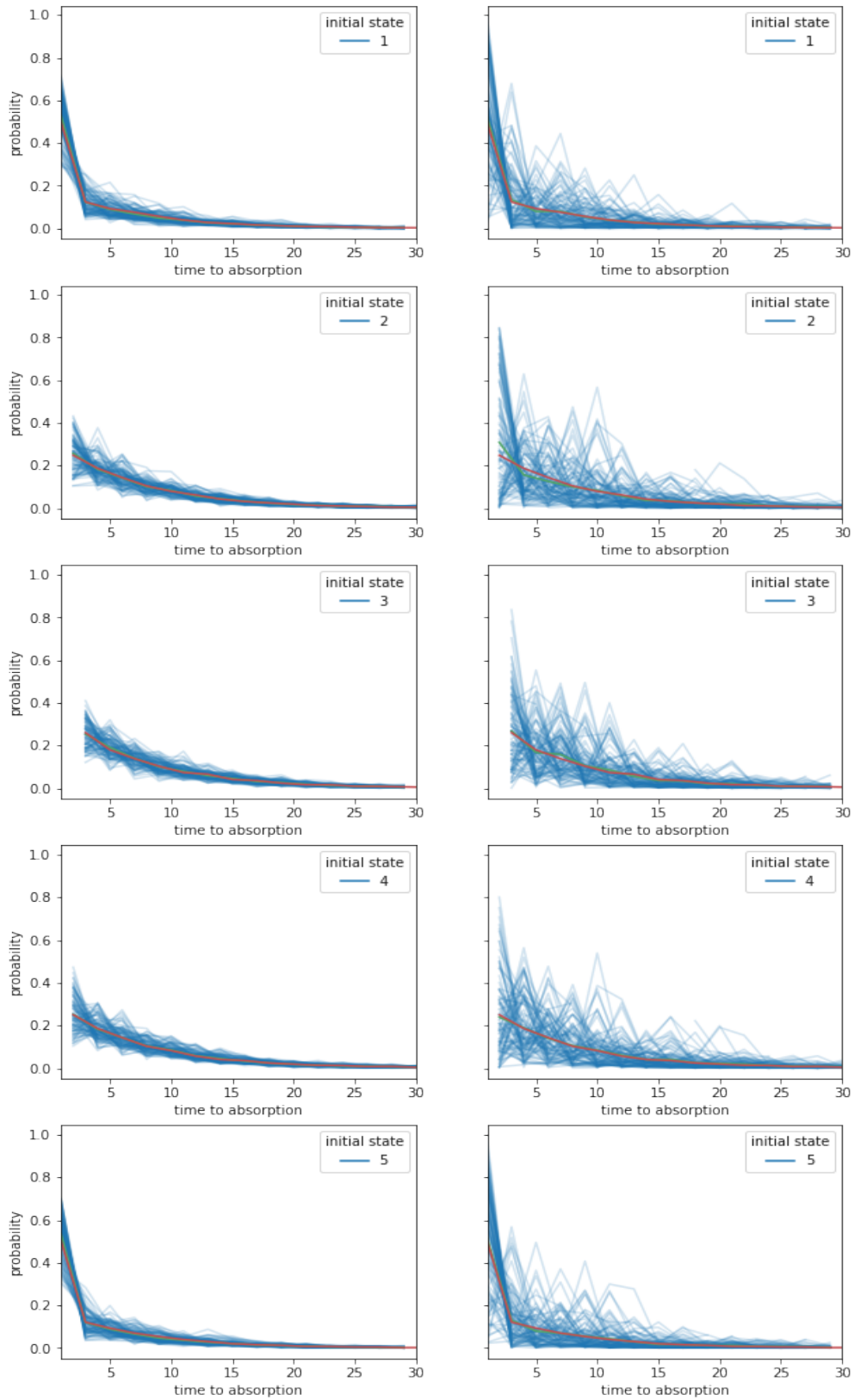
Figure 6.1.8: Histograms of the time to absorption of simulations of the perturbed gambler's ruin with $\varepsilon = 0.2$ (left) and 0.5 (right)

6.2 Moran Processes with Temporal Randomness

6.2.1 Introduction

In this section, we will propose a formulation of Moran processes that allows for time varying randomness. The Moran process model is a classic framework for modeling biological systems with asexual reproduction and a competition for survival between two organisms. It was originally proposed in [24] and has since had a vast presence in the literature. In a Moran process, we assume that we have a fixed population of size N and that during each time step, one organism is chosen to reproduce and another is chosen to die off. In such models, two central quantities of interest are the probability of fixation and the time until absorption. The probability of fixation refers to the probability that one of the two organisms takes over the entire population and the other goes extinct. The time until absorption is the time until one of the two organisms fixates. In [5, 15, 1, 23], Moran processes with spatial randomness are considered with an interest on how spatial randomness impacts the probability of fixation. In our work, we are interested in exploring Moran processes with temporal randomness. In particular, we will explore the case where the temporal randomness is uniform and we will arrive at analytic expressions of the mean probability of fixation and mean time of absorption.

6.2.2 Deterministic Birth and Death Processes

One classic example of absorbing Markov chains that receives a lot of attention in biology applications is a birth and death process in a fixed population size with only two cells that reproduce asexually and are competing for survival. Consider a fixed population of size N with two cells A and B. Let X_t denote the size of the cell A population and let λ_i and μ_i denote the probabilities of birth and death of cell A. Then the state space of X_t is $\mathcal{S} = \{0, 1, \ldots, N\}$ and the transition matrix of X_t will have two absorbing states and will be

of the form

$$
P = \begin{bmatrix}
1 & & & & & & \\
\mu_1 & 1 - \mu_1 - \lambda_1 & \lambda_1 & & & & \\
& \ddots & & \ddots & & \ddots & \\
& & & \mu_{N-1} & 1 - \mu_{N-1} - \lambda_{N-1} & \lambda_{N-1} \\
& & & & & 1
\end{bmatrix}.
$$

During each time interval, t to $t+1$, the population of cell A either increases by 1, decreases by 1 or stays the same. This is a reasonable assumption only if the time step is sufficiently small. The absorbing states 0 and N capture the possible scenarios of cell A going extinct and cell B going extinct, respectively. Absorption will take place with probability 1. Let ρ_i denote the probability of absorption of the process in state N given initial state i, i.e that cell A with initial population i takes over the whole population and cell B goes extinct. Then $\rho_i = P_{iN}^\infty$, the (i, N)-th entry of $P^\infty := \lim_{t \to \infty} P^n$. Let τ_i denote the mean time until absorption of the process given initial state i. Then $\tau_i = \left[(I - B)^{-1} \mathbf{1} \right]_i$ for $i = 1, \ldots, n-1$, where

$$
B = \begin{bmatrix}
1 - \mu_1 - \lambda_1 & \lambda_1 & & & \\
& \ddots & & \ddots & \\
& & \mu_{N-1} & 1 - \mu_{N-1} - \lambda_{N-1}
\end{bmatrix}.
$$

6.2.3   Birth and Death Processes with Temporal Randomness

In this section, our goal is to generalize and study birth and death processes with temporal randomness. Temporal randomness refers to randomness that is time dependent, as opposed to spatial randomness which models environmental randomness that is independent of time. The two primary quantities of interest in this study are the probability of absorption in state N and the mean time until absorption. Consider a birth and death process as presented above with λ_i and μ_i random variables. Then X_t is an absorbing Markov chain with state space

$\mathcal{S} = \{0, 1, \ldots, N\}$ and random transition matrix

$$P(\omega) = \begin{bmatrix} 1 & & & & & \\ \mu_1(\omega) & 1 - \mu_1(\omega) - \lambda_1(\omega) & \lambda_1(\omega) & & & \\ & \ddots & \ddots & \ddots & & \\ & & \mu_{N-1}(\omega) & 1 - \mu_{N-1}(\omega) - \lambda_{N-1}(\omega) & \lambda_{N-1}(\omega) \\ & & & & 1 \end{bmatrix}.$$

where $\omega \in \Omega$ and Ω is the set of all possible outcomes. Assume $P(\omega)$ has mean

$$P_0 = \begin{bmatrix} 1 & & & & \\ \mu_1^0 & 1 - \mu_1^0 - \lambda_1^0 & \lambda_1^0 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_{N-1}^0 & 1 - \mu_{N-1}^0 - \lambda_{N-1}^0 & \lambda_{N-1}^0 \\ & & & & 1 \end{bmatrix}.$$

Let

$$P^{(n)}(\overline{\omega}) := \prod_{i=1}^{n} P(\omega_i)$$

be the n-product of $P(\omega_i)$ iid for $i = 1, 2, \ldots, n$ and $\overline{\omega} = (\omega_1, \omega_2, \ldots) \in \Omega^\infty$. Let

$$P^{(\infty)}(\overline{\omega}) := \lim_{n \to \infty} P^{(n)}(\overline{\omega}).$$

We know that $P^{(n)}(\overline{\omega})$ converges almost surely in n by our previous theoretical results. By the independence assumption,

$$\mathbb{E}\left[P^{(\infty)}(\overline{\omega})\right] = \prod_{i=1}^{\infty} P_0 =: P_0^\infty.$$

Let $\rho_i(\overline{\omega})$ denote the probability of absorption in state N given initial state i in the random case. Let ρ_i^0 denote the same quantity in the non-random case with resect to the Markov

chain with transition matrix P_0. Then

$$\mathbb{E}\left[\rho_i\left(\overline{\omega}\right)\right] = \rho_i^0.$$

Thus we see that the probability of absorption of the random process coincides with the probability of absorption of the deterministic process driven by the mean of the random transition matrix.

Next, we will explore the mean time until absorption. In the case of a classic absorbing chain with transition matrix

$$P = \begin{bmatrix} I & O \\ R & Q \end{bmatrix},$$

the fundamental matrix is given by $F := I + Q + Q^2 + \cdots = (I - Q)^{-1}$ and the expected time until absorption is given by $F\mathbf{1}$. In the context of inhomogeneous absorbing Markov chains driven by the sequence of matrices $P_1, P_2, \ldots$, the time until absorption is according to

$$F = I + \sum_{m=1}^{\infty}\prod_{k=1}^{m} Q_k = I + Q_1 + Q_1 Q_2 + \cdots .$$

Now, in our random process context, first let

$$P\left(\omega\right) = \begin{bmatrix} 1 & & \\ A\left(\omega\right) & B\left(\omega\right) & C\left(\omega\right) \\ & & 1 \end{bmatrix}$$

and

$$P_0 = \begin{bmatrix} 1 & & \\ A_0 & B_0 & C_0 \\ & & 1 \end{bmatrix}.$$

Then the fundamental matrix of a Markov chain with random transition matrices $P\left(\omega\right)$ is

random and given by

$$F(\overline{\omega}) = I + \sum_{m=1}^{\infty} \prod_{k=1}^{m} B(\omega_i) = I + B(\omega_1) + B(\omega_1) B(\omega_2) + \cdots \qquad \text{for } \overline{\omega} \in \Omega^{\infty}.$$

So

$$\mathbb{E}\left[F(\overline{\omega})\right] = \sum_{m=0}^{\infty} B_0^m = (I - B_0)^{-1} =: F_0$$

by the independence assumption. So the mean time until absorption is given by

$$\mathbb{E}\left[F(\overline{\omega}) \mathbf{1}\right] = F_0 \mathbf{1}.$$

Let τ_i denote the mean time until absorption given initial state i. Then τ_i is the i-th entry of the column vector $F_0 \mathbf{1}$. Again, we see that the mean time until absorption of the random process coincides with the mean time until absorption of the deterministic process driven by the mean of the random transition matrix. Thus, if we have the mean of $P(\omega)$ then we can analytically arrive at the mean probability of absorption in state N and the mean time until absorption.

6.2.4 Analytic Expression for the Probability of Absorption in State N

In the above discussion, we are able to analytically calculate the mean probability of absorption in state N and the mean time until absorption by means of an inverse matrix or infinite sum of matrices. In this section, we will further develop these analytic results to be comprised of finite sums and products and only in terms of the birth and death probabilities λ_i and μ_i. Moreover, these results will allow us to calculate the mean probabilities of absorption in the random case.

Consider X_t an absorbing Markov chain with state space $\mathcal{S} = \{0, 1, \ldots, N\}$ and random

transition matrix

$$
P = \begin{bmatrix}
1 & & & & & \\
\mu_1 & 1-\mu_1-\lambda_1 & \lambda_1 & & & \\
& \ddots & \ddots & & \ddots & \\
& & \mu_{N-1} & 1-\mu_{N-1}-\lambda_{N-1} & \lambda_{N-1} \\
& & & & 1
\end{bmatrix}.
$$

We will index the rows and columns of P from 0 to N. Let B be the sub-matrix of dimension $(N-1) \times (N-1)$ not containing the absorbing states. Then the probability of absorption in state N given initial state i is

$$
\begin{aligned}
\rho_i &= \left(I + B + B^2 + \cdots\right)_{i,N-1} P_{N-1,N} \\
&= (I-B)^{-1}_{i,N-1} P_{N-1,N}.
\end{aligned}
$$

Let $D = I - B$. Then

$$
D = \begin{bmatrix}
\mu_1+\lambda_1 & -\lambda_1 & & & & \\
-\mu_2 & \mu_2+\lambda_2 & -\lambda_2 & & & \\
& \ddots & \ddots & & \ddots & \\
& & -\mu_{N-2} & \mu_{N-2}+\lambda_{N-2} & -\lambda_{N-2} \\
& & & -\mu_{N-1} & \mu_{N-1}+\lambda_{N-1}
\end{bmatrix}.
$$

Due to the tridiagonal form of D, its inverse can be explicitly calculated.

The inverse of an n-dimensional tridiagonal matrix of the from

$$
T = \begin{bmatrix}
a_1 & b_1 & & & & \\
c_1 & a_2 & b_2 & & & \\
& c_2 & \ddots & \ddots & & \\
& & \ddots & \ddots & b_{n-1} \\
& & & c_{n-1} & a_n
\end{bmatrix}
$$

is given by

$$
\left(T^{-1}\right)_{ij} = \begin{cases}
(-1)^{i+j}\, b_i \cdots b_{j-1}\theta_{i-1}\phi_{j+1}/\theta_n & \text{if } i < j \\[2mm]
\theta_{i-1}\phi_{j+1}/\theta_n & \text{if } i = j \\[2mm]
(-1)^{i+j}\, c_j \cdots c_{i-1}\theta_{j-1}\phi_{i+1}/\theta_n & \text{if } i > j
\end{cases}
$$

where θ_i satisfy

$$
\theta_i = a_i\theta_{i-1} - b_{i-1}c_{i-1}\theta_{i-2} \quad \text{for } i = 2, 3, \ldots, n
$$

with $\theta_0 = 1, \theta_1 = a_1$ and ϕ_i satisfy

$$
\phi_i = a_i\phi_{i+1} - b_i c_i \phi_{i+2} \quad \text{for } i = n-1, \ldots, 1
$$

with $\phi_{n+1} = 1$ and $\phi_n = a_n$. [36]

In our context, with $a_i = \mu_i + \lambda_i$, $b_i = -\lambda_i$ and $c_i = -\mu_{i+1}$ we get

$$
\theta_\ell = \prod_{i=1}^{\ell} \mu_i + \sum_{k=1}^{\ell-1}\left[\left(\prod_{i=1}^{k}\mu_i\right)\left(\prod_{j=k+1}^{\ell}\lambda_j\right)\right] + \prod_{j=1}^{\ell}\lambda_j \quad \text{for } \ell = 2, 3, \ldots, N-1
$$

with $\theta_0 = 1, \theta_1 = \mu_1 + \lambda_1$ and

$$
\phi_\ell = \prod_{i=\ell}^{N-1}\mu_i + \sum_{k=1}^{N-1-\ell}\left[\left(\prod_{i=\ell}^{N-1-k}\mu_i\right)\left(\prod_{j=N-k}^{N-1}\lambda_j\right)\right] + \prod_{j=\ell}^{N-1}\lambda_j \quad \text{for } \ell = N-2, \ldots, 1
$$

with $\phi_N = 1$ and $\phi_{N-1} = \mu_{N-1} + \lambda_{N-1}$. So

$$\rho_\ell = D_{\ell,N-1}^{-1} P_{N-1,N}$$

$$= \frac{\displaystyle\sum_{k=1}^{\ell-1}\left[\left(\prod_{i=1}^{k}\mu_i\right)\left(\prod_{j=k+1}^{N-1}\lambda_j\right)\right] + \prod_{j=1}^{N-1}\lambda_j}{\displaystyle\prod_{i=1}^{N-1}\mu_i + \sum_{k=1}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i\right)\left(\prod_{j=k+1}^{N-1}\lambda_j\right)\right] + \prod_{j=1}^{N-1}\lambda_j}$$

$$= \frac{\displaystyle\sum_{k=1}^{\ell-1}\left(\prod_{i=1}^{k}\frac{\mu_i}{\lambda_i}\right) + 1}{\displaystyle\sum_{k=1}^{N-1}\left(\prod_{i=1}^{k}\frac{\mu_i}{\lambda_i}\right) + 1}.$$

Thus ρ_ℓ is increasing with respect to ℓ.

6.2.5   Moran Processes with Temporal Randomness

In this section we will explore one particular example of a birth and death process: a Moran process. A Moran process is a birth and death process with birth and death probabilities

$$\lambda_i = \frac{f_i \cdot i}{f_i \cdot i + g_i(N-i)} \cdot \frac{N-i}{N} \quad \text{and} \quad \mu_i = \frac{g_i(N-i)}{f_i \cdot i + g_i(N-i)} \cdot \frac{i}{N},$$

where f_i and g_i denote the division rates of the two cells A and B. The framework that we have established allows us to consider Moran processes with a wide range of different behaviors depending on the context.

As an example, consider a Moran process with uniform temporal randomness introduced by taking the division rates f_i and g_i to be uniformly distributed on the interval $(1 - \varepsilon, 1 + \varepsilon)$ for some small $\varepsilon \in (0, 1)$. Then the mean of λ_i and μ_i can be calculated by the following lemma.

Lemma 6.2.1. *Let* $X(a, b, c) = \frac{c}{c+1} \cdot \frac{u}{cv+u}$ *with* u, v *uniformly distributed on the interval*

(a, b) *and* c *constant with* $a, b, c > 0$. *Then*

$$\mathbb{E}[X] = \frac{1}{2(b-a)^2(c+1)} \ln\left[\left(\frac{b(c+1)}{ac+b}\right)^{b^2}\left(\frac{a(c+1)}{bc+a}\right)^{a^2}\left(\frac{ac+b}{a(c+1)}\right)^{a^2c^2}\left(\frac{bc+a}{b(c+1)}\right)^{b^2c^2}\right]$$
$$+ \frac{c}{2(c+1)}.$$

The calculation of the mean is simple and merely involves calculus integration. We will include the proof for completeness.

Proof. Note,

$$\mathbb{E}\left[\frac{u}{cv+u}\right] = \frac{1}{(b-a)^2}\int_a^b\int_a^b \frac{u}{cw+u}dwdu = \frac{1}{(b-a)^2}\int_a^b\left[\frac{u}{c}\ln(cw+u)\right]_a^b du$$

$$= \frac{1}{(b-a)^2}\int_a^b \frac{u}{c}\ln\left(\frac{bc+u}{ac+u}\right)du = \frac{1}{(b-a)^2}\int_a^b \ln\left(\frac{bc+u}{ac+u}\right)d\left(\frac{u^2}{2c}\right)$$

$$= \frac{1}{2(b-a)^2}\left[\frac{u^2}{c}\ln\left(\frac{bc+u}{ac+u}\right) - \int \frac{u^2(a-b)}{(ac+u)(bc+u)}du\right]_a^b$$

$$= \frac{1}{2(b-a)^2}\left[\frac{u^2}{c}\ln\left(\frac{bc+u}{ac+u}\right) - \int\left(-\frac{a^2c}{ac+u} + \frac{b^2c}{bc+u} + (a-b)\right)du\right]_a^b$$

$$= \frac{1}{2(b-a)^2}\left[\frac{u^2}{c}\ln\left(\frac{bc+u}{ac+u}\right) + a^2c\ln(ac+u) - b^2c\ln(bc+u) - (a-b)u\right]_a^b$$

$$= \frac{1}{2(b-a)^2}\left[\frac{b^2}{c}\ln\left(\frac{bc+b}{ac+b}\right) + \frac{a^2}{c}\ln\left(\frac{ac+a}{bc+a}\right) + a^2c\ln\left(\frac{b+ac}{a+ac}\right)\right.$$
$$\left. + b^2c\ln\left(\frac{a+bc}{b+bc}\right) + (a-b)^2\right].$$

$\square$

Note that in a Moran process with uniform temporal randomness,

$$\lambda_i = X\left(1-\varepsilon, 1+\varepsilon, \frac{N-i}{i}\right) \quad \text{and} \quad \mu_i = X\left(1-\varepsilon, 1+\varepsilon, \frac{i}{N-i}\right).$$

With this lemma we are able to analytically calculate probability of absorption in state N and the mean time until absorption. We calculated these quantities for $N = 10, 25$ and 50

and included in the figure below.

In the plots below we see a symmetry in the mean probability of absorption. We will formulate and prove this symmetry in the following proposition.

Proposition 6.2.1. *In a Moran process with uniform temporal randomness, we have*

$$\rho_\ell + \rho_{N-\ell} = 1,$$

where ρ_ℓ denotes the mean probability of absorption given initial state ℓ.

Proof. Let

$$\lambda_i^0 = \mathbb{E}\left[X\left(1-\varepsilon, 1+\varepsilon, \frac{N-i}{i}\right)\right]$$

and

$$\mu_i^0 = \mathbb{E}\left[X\left(1-\varepsilon, 1+\varepsilon, \frac{i}{N-i}\right)\right].$$

Then the mean probability of absorption is given by

$$\rho_\ell = \frac{\sum_{k=1}^{\ell-1}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0}{\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=1}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0},$$
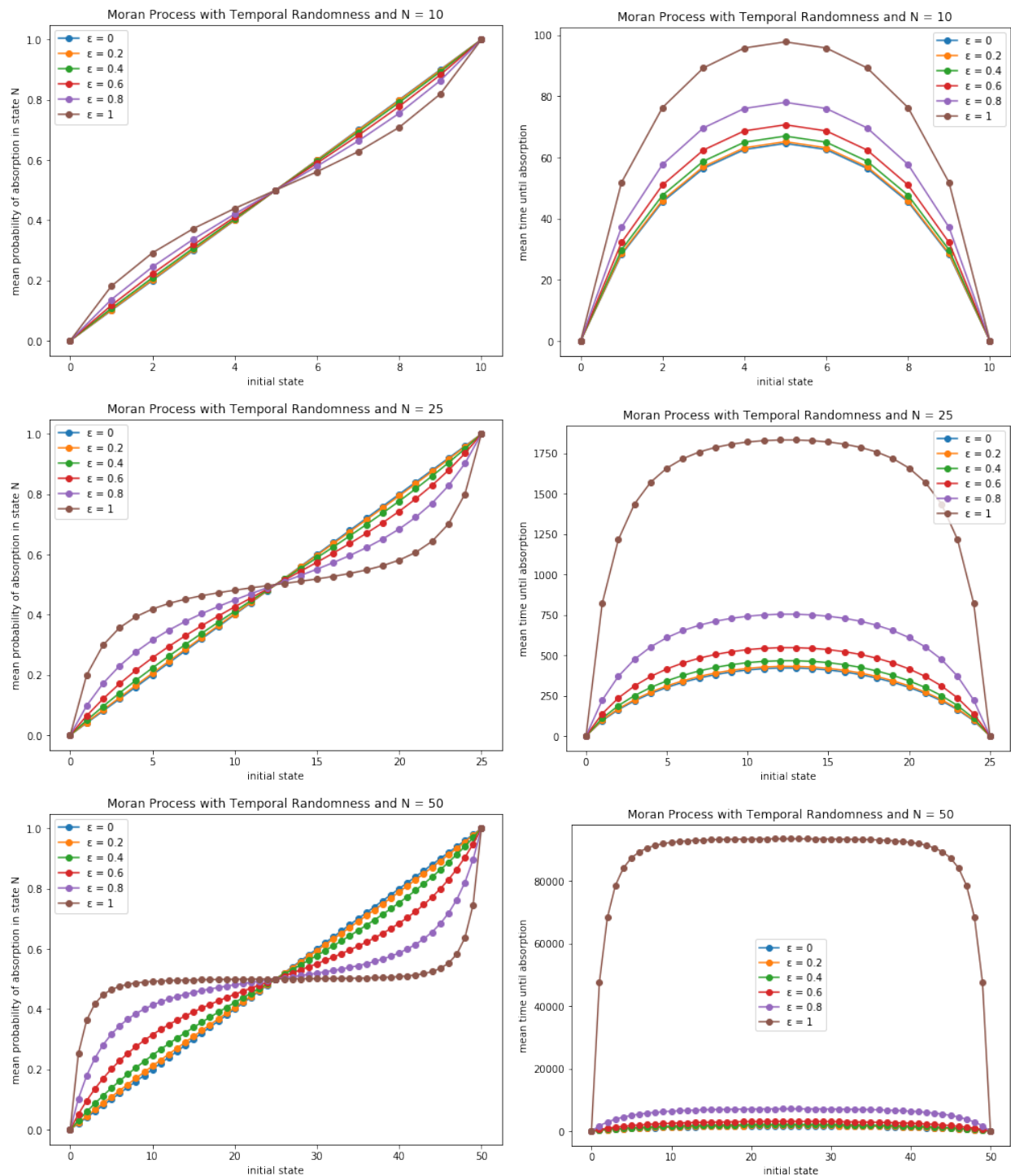
Figure 6.2.1: Mean probability of absorption in state N and the mean time until absorption.

as shown above. Note that $\mu_{N-i}^0 = \lambda_i^0$. So,

$$
\rho_\ell = \frac{\displaystyle\sum_{k=1}^{\ell-1}\left[\left(\prod_{i=1}^{k}\lambda_{N-i}^0\right)\left(\prod_{j=k+1}^{N-1}\mu_{N-j}^0\right)\right] + \prod_{j=1}^{N-1}\mu_{N-j}^0}{\displaystyle\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=1}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0}
$$

$$
= \frac{\displaystyle\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=1}^{\ell-1}\left[\left(\prod_{i=1}^{N-k-1}\mu_i^0\right)\left(\prod_{j=N-k}^{N-1}\lambda_j^0\right)\right]}{\displaystyle\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=1}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0}
$$

$$
= \frac{\displaystyle\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=N-\ell-2}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right]}{\displaystyle\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=1}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0}
$$

$$
= 1 - \frac{\displaystyle\sum_{k=1}^{N-\ell-1}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0}{\displaystyle\prod_{i=1}^{N-1}\mu_i^0 + \sum_{k=1}^{N-2}\left[\left(\prod_{i=1}^{k}\mu_i^0\right)\left(\prod_{j=k+1}^{N-1}\lambda_j^0\right)\right] + \prod_{j=1}^{N-1}\lambda_j^0} = 1 - \rho_{N-\ell}.
$$

$\square$

Chapter 7

Conclusions and Future Work

In this thesis we have motivated and presented Markov chains with random transition matrices with an emphasis on absorbing chains and their applications. This first application of such processes that we discussed was the credit risk model which incorporated the dependence of credit migrations on the underlying economy. This was used to capture the risk inherent in a portfolio of loans. We then presented an overview of Markov chains with random transition matrices and the relevant work in the literature for processes with transition matrices having simple eigenvalue 1. In Proposition 4.3.1, we approximated the asymptotic fluctuation that takes place for such processes within the context of perturbed Markov chains. We also presented a procedure to simulate such chains. In Proposition 5.2.1, we motivated and proved the almost sure convergence of infinite products of random transition matrices with absorbing states. We also approximated the variance of the time until absorption for perturbed absorbing Markov chains. We then studied Moran processes. With our framework, we introduced a form of temporal randomness to Moran processes and derived an analytic result for the probability of fixation. While this work answered interesting and relevant questions, it has also given rise to other questions.

In the context of Markov chains with random transition matrices several questions came up that we where not able to answer, but would provide more insight. For chains with simple

eigenvalue 1, we found that the entries of the $P^{(n)}$ fluctuated about the invariant distribution of $\mathbb{E}\left[P\right]$. One interesting question here is whether we can bound this fluctuation. We also assumed that the transition matrices where independent. An interesting question is what happens when we relax this assumption? Relaxing this assumption can be done in many ways. One way of doing so, which has been studied by [11, 9, 10, 12, 13, 6], is to impose a Markov structure on the switching between the transition matrices. However, this has only been done with irreducible transition matrices. An interesting study would be to take the underling Markov structure to be absorbing.

In the Gambler's ruin application, we studied the case where the probability of wins and loses is uniformly distributed. From our simulations, we came up with Conjecture 6.1.1, which remains unproven.

In the Moran process application, we studied the case where the temporal randomness is modeled by a uniform distribution. One phenomenon that we believe to be happening is that the probability of absorption $\rho_1\left(N, \varepsilon\right)$ converges to a non-zero value as N tends to infinity (at least for some values of ε). If this is true, this is a surprising result because for the classic Moran process, $\rho_1 = \frac{1}{N}$ and converges to 0. The interpretation of such a potential result is that by introducing temporal randomness to the process, the possibility of fixation of a mutant cell will always be non-negligible regardless of how small the initial population of the mutant cell is or how large the overall population N is.

Our work in this thesis has addressed several important questions within the study of Markov models, from theoretical developments to applications in finance and biology.

Bibliography

[1] A. D. Long N. L. Komarova M. Kohandel A. Mahdipour-Shirayeh, A. H. Darooneh. Genotype by random environmental interactions gives an advantage to non-favored minor alleles. *Scientific Reports*, 7(5193):1–8, 2017.

[2] Krishna B. Athreya and Samuel Karlin. Branching processes with random environments, ii: Limit theorems. *The Annals of Mathematical Statistics*, 42(6):1843–1858, 1971.

[3] Krishna B. Athreya and Samuel Karlin. On branching processes with random environments: I: Extinction probabilities. *The Annals of Mathematical Statistics*, 42(5):1499–1520, 1971.

[4] Anil Bangia, Francis X. Diebold, Andre Kronimus, Christian Schagen, and Til Schuermann. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26:445–474, 2002.

[5] Dmitri V. Boreiko, Yuri M. Kaniovski, and Georg Ch. Pflug. *Modeling of Dependent Credit Rating Transitions Governed by Industry-Specific Markovian Matrices*, pages 525–531. 2017.

[6] Richard D. Bourgin and Robert Cogburn. On determining absorption probabilities for markov chains in random environments. *Advances in Applied Probability*, 13(2):369–387, 1981.

[7] Laurent Bruneau, Alain Joye, and Marco Merkli. Infinite products of random matrices and repeated interaction dynamics. *Annales de l'Institut Henri Poincare - Probabilites et Statistiques*, 46(2):442–464, 2010.

[8] Florentin Butaru, Qingqing Chen, Brian Clark, Sanmay Das, Andrew W. Lo, and Akhtar Siddique. Risk and risk management in the credit card industry. *Journal of Banking and Finance*, 72:218–239, 2016.

[9] Robert Cogburn. The ergodic theory of markov chains in random environments. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (1):109–128.

[10] Robert Cogburn. On products of random stochastic matrices. *American Mathematical Society*, (86):199–213.

[11] Robert Cogburn. Markov chains in random environments: The case of markovian environments. *The Annals of Probability*, 8(5):908–916, 1980.

[12] Robert Cogburn. On direct convergence and periodicity for transition probabilities of markov chains in random environments. *The Annals of Probability*, 18(2):642–654, 1990.

[13] Robert Cogburn. On the central limit theorem for markov chains in random environments. *The Annals of Probability*, 19(2):587–604, 1991.

[14] Guglielmo D'Amico, Jacques Janssen, and Raimondo Manca. Downward migration credit risk problem: a non-homogenous backward semi-markov reliability approach. *Journal of the Operational Research Society*, 67(3):393–401, 2016.

[15] Suzan Farhang-Sardroodi, Amir H. Darooneh, Moladad Nikbakht, Natalia L. Komarova, and Mohammad Kohandel. The effect of spatial randomness on the average fixation time of mutants. *PLOS Computational Biology*, 13(11):1–22, 11 2017.

[16] Matthew Fitzpatrick and Dobrin Marchev. Efficient bayesian estimation of the multivariate double chain markov model. *Statistics and Computing*, 23(4):467–480, 2013.

[17] Matthew Anthony Fitzpatrick. Multi-regime models involving markov chains. 2016.

[18] Halina Frydman and Til Schuermann. Credit rating dynamics and markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075, 2008.

[19] Robert G. Gallager. *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.

[20] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability*. American Mathematical Society, 2012.

[21] Robert A. Jarrow, David Lando, and Stuart M. Turnbull. A markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2):481–523, 1997.

[22] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer-Verlag, 1976.

[23] V. S. K. Manem, K. Kaveh, M. Kohandel, and S. Sivaloganathan. Modeling invasion dynamics with spatial random-fitness due to micro-environment. *PLOS ONE*, 10(10):1–20, 10 2015.

[24] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60?71, 1958.

[25] Pamela Nickell, William Perraudin, and Simone Varotto. Stability of rating transitions. *Journal of Banking & Finance*, 24(1-2):203–227, 2000.

[26] Federal Reserve. Consumer credit - g.19, 2017.

[27] L. Saloff-Coste and J. Zuniga. Convergence of some time inhomogeneous markov chains via spectral techniques. *Stochastic Processes and their Applications*, 117:961–979, 2007.

[28] L. Saloff-Coste and J. Zuniga. Merging for time inhomogeneous finite markov chains, part i: singular values and stability. *Electronic Journal of Probability*, 14(49):1456–1494, 2009.

[29] L. Saloff-Coste and J. Zuniga. Time inhomogeneous markov chains with wave-like behavior. *The Annals of Applied Probability*, 20(5):1831–1853, 2010.

[30] L. Saloff-Coste and J. Zuniga. Merging and stability for time inhomogeneous finite markov chains. 2013.

[31] Walter L. Smith. Necessary conditions for almost sure extinction of a branching process with random environment. *Ann. Math. Statist.*, 39(6):2136–2140, 12 1968.

[32] Walter L. Smith and William E. Wilkinson. On branching processes in random environments. *Ann. Math. Statist.*, 40(3):814–827, 06 1969.

[33] Yukio Takahashi. Markov chains with random transition matrices. *Kodai Mathematical Journal*, 21:426–447, July 1969.

[34] William C. Torrez. The birth and death chain in a random environment: Instability and extinction theorems. *The Annals of Probability*, 6(6):1026–1043, 1978.

[35] William C. Torrez. Calculating extinction probabilities for the birth and death chain in a random environment. *Journal of Applied Probability*, 16(4):709–720, 1979.

[36] Riaz A. Usmani. Inversion of a tridiagonal jacobi matrix. *Linear Algebra and its Applications*, 212-213:413 – 414, 1994.

[37] William E. Wilkinson. On calculating extinction probabilities for branching processes in random environments. *Journal of Applied Probability*, 6(3):478–492, 1969.