Path Properties of Rare Events

Jesse Collingwood

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies
In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

© Jesse Collingwood, Ottawa, Canada, 2015

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

Simulation of rare events can be costly with respect to time and computational resources. For certain processes it may be more efficient to begin at the rare event and simulate a kind of reversal of the process. This approach is particularly well suited to reversible Markov processes, but holds much more generally. This more general result is formulated precisely in the language of stationary point processes, proven, and applied to some examples.

An interesting question is whether this technique can be applied to Markov processes which are substochastic, i.e. processes which may *die* if a graveyard state is ever reached. First, some of the theory of substochastic processes is developed; in particular a slightly surprising result about the rate of convergence of the distribution π_n at time n of the process conditioned to stay alive to the quasi-stationary distribution, or Yaglom limit, is proved. This result is then verified with some illustrative examples.

Next, it is demonstrated with an explicit example that on infinite state spaces the reversal approach to analyzing both the rate of convergence to the Yaglom limit and the likely path of rare events can fail due to transience.

Acknowledgements

Many people have been gracious guides during the writing of my thesis. Jean-Yves Le Boudec kindly provided me with an office to work in and bits of lunchtime wisdom at École Polytechnique Fédérale de Lausanne, and it was here that I met with the invaluable insight of Nicolas Gast whose help in obtaining some of my results cannot be overstated. Gail Ivanoff offered encouraging words of support on more than one occasion; for this and for the years of helpful tutelage she imparted to me before this project ever began I am very grateful. Paul Dupuis must be sincerely thanked for his help in overcoming a sticky technical issue. I am indebted to Chris Dionne and Elizabeth Maltais for their support as friends and colleagues, and to several of the graduate students with whom I have had the pleasure of sharing close academic quarters over the past 4 years, among them Farid Elktaibi, Maryam Sohrabi, Nada Habli, Ibrahim Abdelrazeq, and Jeanseong Park.

As with anyone, the most vital support I received was at home, and consequently my deepest gratitude lies with my own family and with the Lee-Shanok family. If this thesis is a play they are its light crew, make up artist, special effects technician and one hundred other things that work behind the scenes to make everything possible.

Finally, I must thank my friend and advisor David McDonald. What I have learned under his tutelage intersects many domains from mathematics to history and politics to where one finds the best eats in Paris. I have learned to take a more intuitive approach to things from being exposed to his thinking style, and I am eternally grateful for the opportunities he has furnished me with.

Dedication

For Murray.

Contents

Chapter 1

Introduction

There are two central ideas upon which this thesis is based. The first is the fact, known already to some, that when attempting to understand what a large deviation path for a process looks like the prudent course may be to begin at the end and look at things in reverse. This is because straightforward simulation may take forever to yield the desired rare event; however, if the laws governing the time reversal are available then by beginning at the rare event and "running the film" of its path backwards one selects from the very subset of the forward paths which are of interest.

This general idea is not new, and has obscure origins ([2] is the earliest example of which I am aware, and was never published). Yet like so many folk tales and ballads which come through mysterious beginnings to fixate in our minds, it is a part of the repertoire of many probabilists. It is for this reason that the result has been named the Folk Theorem herein. It has not to my knowledge been given a rigorous formulation and proof for point processes before now. This is done in Section 2.4 after first setting up the notation and framework in Sections 2.1 through 2.3.

Section 2.1 follows closely the standard text [3] and introduces the reader to the theory of stationary point processes. The purpose of Section 2.2 is to introduce processes *watched* on a set and to arrive at Lemma 2.2.2, which is a new calculation based on a similar formula in [3]. This formula provides a means of calculating certain kinds of Palm probabilities from other known quantities and is subsequently used to help apply the Folk Theorem.

Section 2.3 deals with reversals. The definition of a reversal of a sample path and a stochastic process is given, and all the changes to the corresponding stationary point process are carefully tracked. The main facts needed from this Section are Lemma 2.3.3, which simply says that reversing a set of paths does not change its probability under the stationary measure, and Equation 2.3.1, which is a way of rewriting the reversal of a set shifted by a particular exit time and is needed in the proof of the Folk Theorem.

In Section 2.4 a simple coin flipping example is given to illustrate the central idea behind the Folk Theorem, and then the Theorem itself is presented. Its rigorous formulation and proof using the theory of stationary point processes is one of the main results of the thesis. The goal is to apply the Folk Theorem to Markov chains, and so the subsequent subsection provides techniques for computing the Palm probability expressions of the Theorem from the stationary measure and kernel of a given chain. Lemmas 2.4.3 and 2.4.4 are new results which achieve this, and Corollary 2.4.5 uses these two Lemmas to essentially restate the Folk Theorem in the language of Markov chains. Section 2.4.3 applies the Markovian version of the Folk Theorem to an accounting network example taken from [8].

The next idea, also already well known, is that the standard paradigm wherein one approximates the distribution of a Markovian process at time n, for large n, by the limiting steady state may be an oversimplification in some cases. This is because many processes so modeled are in fact substochastic - i.e. there is a chance that the whole process stops (the machine breaks, the population dies, the universe implodes etc). For such processes analyzing the distribution at time n *conditional upon still being alive* and its limiting distribution π (when it has one) may be more apt. Thus the main result of Chapter 3 gives a rate of convergence of these conditional distributions to π for discrete time Markov chains. Notice that a substochastic chain staying alive until a large time is itself a rare event.

To set up the exposition, Section 3.1 introduces the concept of a Yaglom limit and states some of the classic results for such objects. Then Theorem 3.3.2 of Section 3.3 is the central new result giving the rate of convergence of π_n to π with respect to a chi-square distance. The examples to which this result is then applied (in Section 3.4) illustrate the surprising fact that one of the standard approaches to finding such rates of convergence in the fully stochastic case (given in [11]) fails in the substochastic case. Another important consideration in the substochastic setting is whether the convergence to quasi-stationarity is more rapid than absorption, and in Section 3.4 a simple eigenvalue inequality is given as a test and applied in an example. Section 3.5 restates the ideas of Section 3.3 in the context of continuous time chains.

Bringing these two ideas together to analyze large deviations of substochastic chains is discussed in Chapter 4. The central idea of this chapter is that large deviations require many time steps to occur, and thus the kernel governing the transitions for such paths is the one formed by conditioning on being alive not at the present, but far into the future. In the limit, one considers the *sustained kernel* of Section 4.1, whose paths are conditioned to live indefinitely. Such paths typically form a set of measure 0 since absorption (i.e. death) is usually assumed to be certain. The measure induced by the sustained kernel appeared in [7].

No rigorous version of the Folk Theorem for substochastic chains is proved. However a new example based on the network of [8] is given in Section 4.2 which illustrates both that considering the sustained kernel is the justified approach, but also that the Folk Theorem can fail to hold. This happens because on infinite state spaces conditioning on never dying can cause the sustained kernel to "push" paths away to infinity in order to keep them alive. The trade off for immortality is exile. Nevertheless, in the example of Section 4.2 the backwards process does adhere to a predictable path for a time, then hits a boundary and flies away. The "fix" for such behavior is to put reflecting bounds around the state space to make it finite.

Finding the quasi-stationary distribution π is often more difficult than finding the steady state of a stochastic Markov kernel, and so finally in Chapter 5 two techniques for constructing the quasi-stationary measure are given. First in Section 5.1 the watched process of Section 2.2 is revisited in the substochastic setting. Here the ideas of Section 4 of [25] are extended to the case where the set A on which the process is being "watched" is more general than a singleton. A new representation Theorem (Theorem 5.1.6) shows how quasi-stationary measures can be computed from a 1-invariant measure of a much smaller matrix indexed by elements of the watched set A.

In Section 5.2 the Approximation Lemma (Lemma 5.2.1) shows how the quasi-stationary measure can be constructed from an approximate quasi-stationary measure (i.e. a measure which satisfies the necessary equations on a reduction of the state space). There are cases in which finding such measures is easier than finding π because parts of the state space on which the transitions are more difficult to work with can be "swept away". This is a new result based on a similar approach given in [8] for stochastic kernels. The Approximation Lemma is applied to an example in which π is successfully computed.

Throughout, probability measures on a countable space S will typically be treated as row vectors when left multiplying matrices: πK. The vectors which contain only 0s or 1s will be treated as column vectors, as will generally any functions to which a matrix will be applied as an operator: Kh. It should be clear from the context which form a function on S takes. For a function g on S the expression $\text{diag}(g)$ is used to denote the square matrix whose rows and columns are indexed by S and which has the entries of g along its diagonal, and 0s elsewhere.

Chapter 2

Rare Excursions for Markov Processes

2.1 Stationary Point Processes

The notation established in [3, 4] is followed, and a brief explanation of it is given in this section.

Let (N, θ_t, P) be a *stationary point process* on the probability space $(\Omega, \mathcal{F}, P)$; this means that

i) N is a random variable on $(\Omega, \mathcal{F}, P)$ taking values in the space $(M, \mathcal{M})$ of sigma finite counting measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with the σ-field $\mathcal{M}$ generated by sets of the form

$$\{m \in M; m(C) = k\}, \quad k \in \mathbb{N}, C \in \mathcal{B}(\mathbb{R}),$$

ii) $\{\theta_t\}_{t \in \mathbb{R}}$ is a measurable flow on $(\Omega, \mathcal{F})$, i.e.

   a) $(t, \omega) \to \theta_t(\omega)$ is $\mathcal{B}(\mathbb{R}) \otimes \mathcal{F}/\mathcal{F}$- measurable,

   b) θ_t is bijective for all $t \in \mathbb{R}$, and

   c) $\theta_t \circ \theta_s = \theta_{t+s}$ for all $t, s \in \mathbb{R}$,

iii) (N, P) is compatible with the flow $\{\theta_t\}_{t\in\mathbb{R}}$, i.e. for every $t \in \mathbb{R}$ and every $\omega \in \Omega$, the following equalities of measures hold: $N(\theta_t\omega) = S_t N(\omega)$ and $P \circ \theta_t = P$, where S_t is the *shift operator* on $(M, \mathcal{M})$ satisfying $S_t m(C) = m(C + t)$.

Note that θ_0 is the identity on Ω and hence that the inverse of θ_t is $\theta_t^{-1} = \theta_{-t}$. The intuitive idea here is that the random measure N and the probability P look no different when shifted by the flow θ_t. The use of the term "stationary" is justified since it follows from these definitions that

$$P[N(C_1) = n_1, \ldots, N(C_k) = n_k] = P[N(C_1 + t) = n_1, \ldots, N(C_k + t) = n_k]$$

for any borel sets $C_1, \ldots, C_k \subseteq \mathbb{R}$, any nonnegative integers $n_1, \ldots, n_k \in \mathbb{N}$, and any $t \in \mathbb{R}$.

The n^{th} point of $N(\omega)$ is denoted $T_n(\omega)$ so that $N(C) = \sum_{n\in\mathbb{Z}} I_C(T_n)$ (the convention here is that the zeroth point of $N(\omega)$ is the largest point not exceeding zero to which $N(\omega)$ gives mass, and that the points are ordered).

A stochastic process $\{X_t\}_{t\in\mathbb{R}}$ on $(\Omega, \mathcal{F}, P)$ is θ_t *compatible* if it satisfies $X_t(\omega) = X_0(\theta_t\omega)$ for all $t \in \mathbb{R}$ and $\omega \in \Omega$. Notice that $P \circ \theta_t = P$ implies that such a process is stationary in the usual sense since

$$P[X_t \in A] = P[X_0 \circ \theta_t \in A] = P \circ \theta_{-t}[X_0 \in A] = P[X_0 \in A]$$

for all $t \in \mathbb{R}$ and all $A \in \mathcal{F}$.

Although taking $(\Omega, \mathcal{F}, P)$ to be any general probability space does not change the results of the discussion herein, more intuitive clarity is achieved (without compromising that generality) by selecting a particular space to deal with. Specifically, let Ω the set of cadlag trajectories $\omega : \mathbb{R} \to S$ from $\mathbb{R}$ to some countable, measurable space $(S, \mathcal{S})$ which have only countably many, non-accumulating discontinuities and let $\mathcal{F} = \sigma(\{X_t^{-1}(G); G \in \mathcal{S}, t \in \mathbb{R}\})$ where $\{X_t\}_{t\in\mathbb{R}}$ is the coordinate process given by $X_t(\omega) = \omega(t)$. According to Kolmogorov's Extension Theorem and Theorem 38.1 of

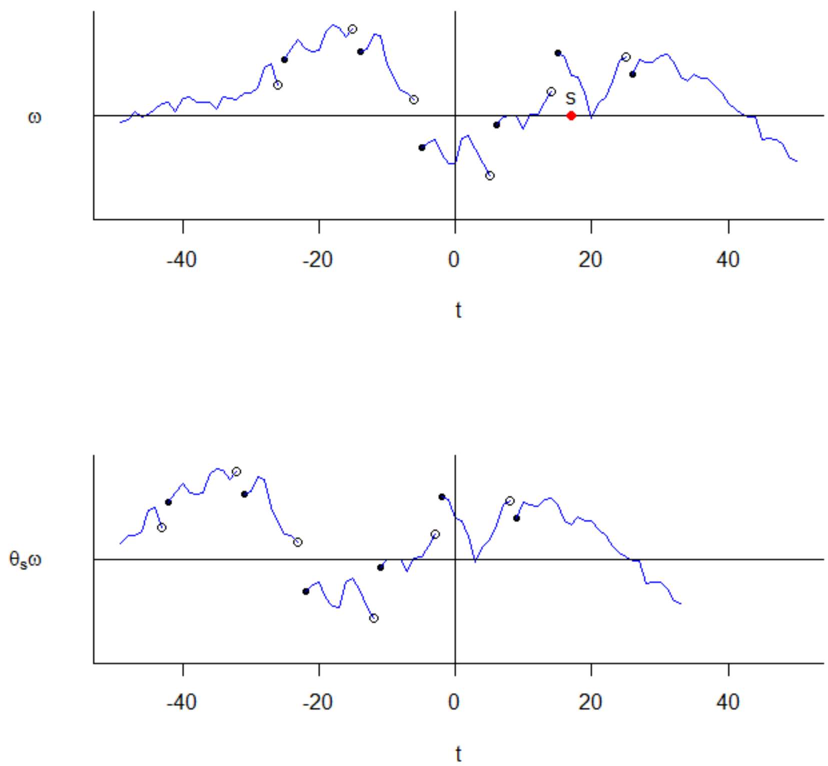

Figure 2.1: A sample path ω and the shifted path $\theta_s\omega$.

[5], it is possible to obtain such a space/process pair corresponding to any collection of finite dimensional distributions one might consistently specify, and this shall be done variously throughout.

Take θ_t to be the shift satisfying $\theta_t\omega(s) = \omega(s+t)$. It is clear that this defines a measurable flow and that the coordinate process is indeed compatible with $\{\theta_t\}_{t\in\mathbb{R}}$. Let P be such that $P \circ \theta_t = P$ for all $t \in \mathbb{R}$ so that the process $\{X_t\}_{t\in\mathbb{R}}$ is a stationary process. Figure 2.1 depicts a typical path ω and the shifted path $\theta_s\omega$. With this setup various point processes can be constructed; for instance, N may count the jump discontinuities of X, or the number of entrances into a particular set.

For any random variable $W : \Omega \to \mathbb{R}$ define $\theta_W : \Omega \to \Omega$ by $\theta_W(\omega) := \theta_{W(\omega)}(\omega)$.

It will be helpful to refer later to the following properties: for random variables $U, V : \Omega \to \mathbb{R}$

$$\theta_U \circ \theta_V = \theta_{V+U\circ\theta_V}. \tag{2.1.1}$$

Also,

$$T_m \circ \theta_t = \begin{cases} T_{m+N(0,t]} - t & t \geq 0 \\ T_{m-N(t,0]} - t & t < 0 \end{cases} \tag{2.1.2}$$

holds for all $m \in \mathbb{Z}$ from which

$$T_m \circ \theta_{T_n} = T_{m+n} - T_n, \quad m, n \in \mathbb{Z} \tag{2.1.3}$$

follows by taking $t = T_n$. Consequently

$$\theta_{T_n} \circ \theta_{T_{-n}} = \theta_{T_{-n}+T_n\circ\theta_{T_{-n}}} = \theta_{T_{-n}+T_{n-n}-T_{-n}} = \theta_{T_0},$$

and so, for all $n \in \mathbb{N}$,

$$\theta_{T_n}^{-1} = \theta_{T_{-n}} \quad \text{on} \quad \Omega_0 := \{T_0 = 0\}. \tag{2.1.4}$$

The focus being on stochastic paths, jump processes on a two dimensional integer lattice for example, it is natural to ask questions like “what is the probability of having hit a distant point $(\ell, y) \in \mathbb{N}^2$ after some large time t?”. In practical examples, such as when the jump process describes the state of a tandem queue, one wishes to do analysis conditional on a jump having occured at time 0. This is a set of measure 0 however, and so the following construction of a measure which puts all its mass on Ω_0 is handy.

Let l denote the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let $\lambda = \mathbb{E}[N(0,1]]$ be the *intensity* of (N, θ_t, P). Since $P \circ \theta_t = P$, the measure $\mathcal{B}(\mathbb{R}) \ni C \mapsto \mathbb{E}[N(C)]$ is translation invariant and therefore proportional to l; that is $\mathbb{E}[N(C)] = \lambda \times l(C)$. Furthermore, for $A \in \mathcal{F}$ and $t \geq 0$, stationarity and (2.1.2) imply that

$$\mathbb{E}\left[\sum_{n\in\mathbb{Z}} I_A(\theta_{T_n}) I_C(T_n)\right] = \mathbb{E}\left[\sum_{n\in\mathbb{Z}} I_A(\theta_{T_n} \circ \theta_t) I_C(T_n \circ \theta_t)\right]$$

$$= \mathbb{E}\left[\sum_{n\in\mathbb{Z}} I_A(\theta_{T_{n+N(0,t]}}) I_C(T_{n+N(0,t]} - t)\right]$$
$$= \mathbb{E}\left[\sum_{k\in\mathbb{Z}} I_A(\theta_{T_k}) I_{C+t}(T_k)\right]$$

with the extremes of this equality holding also for $t \leq 0$ (the argument proceeds with $N(0,t]$ replaced by $-N(t,0]$). That is, the measure

$$\mathcal{B}(\mathbb{R}) \ni C \mapsto \mathbb{E}\left[\sum_{n\in\mathbb{Z}} I_A(\theta_{T_n}) I_C(T_n)\right]$$

is also translation invariant, hence proportional to l, and therefore

$$P_0^N(A) := \frac{1}{\lambda l(C)} \mathbb{E}\left[\sum_{n\in\mathbb{Z}} I_A(\theta_{T_n}) I_C(T_n)\right] = \frac{1}{\lambda l(C)} \mathbb{E}\left[\int_C (I_A \circ \theta_s) N(ds)\right] \quad (2.1.5)$$

has the same value for any $C \in \mathcal{B}(\mathbb{R})$. The resultant measure P_0^N on $\mathcal{F}$ is called the *Palm probability* of the stationary point process (N, θ_t, P). As mentioned earlier, it has the appealing property that

$$P_0^N[\Omega_0] = \frac{1}{\lambda l(C)} \mathbb{E}\left[\sum_{n\in\mathbb{Z}} \chi\{T_0 \circ \theta_{T_n} = 0\} I_C(T_n)\right] = 1 \quad (2.1.6)$$

as $T_0 \circ \theta_{T_n} \equiv 0$.

Korolyuk's estimate ((1.5.1) of [3]) says that

$$P[N(0,t] > 1] = o(t), \quad (2.1.7)$$

and Dobrushin's estimate ((1.5.2) of [3]) says that

$$P[N(0,t] > 0] = \lambda t + o(t);$$

it will be useful to notice that this second limit implies

$$\lim_{t\to 0} \frac{P[T_1 \leq t]}{\lambda t} = 1. \quad (2.1.8)$$

In the forthcoming Folk Theorem stationary point processes which count only jumps of a certain sort, namely entrances or exits into sets of particular interest, shall be studied. Let $A \in \mathcal{S}$ be such that the stochastic process $\{I_{\{X(t)\in A\}}\}_{t\in\mathbb{R}}$ is almost surely continuous at all points t except possibly those belonging to a countable set with no accumulation points.

The discontinuities of the process $\{I_{\{X(t)\in A\}}\}_{t\in\mathbb{R}}$ can be categorized as either *entrance times*, i.e. those values $t \in \mathbb{R}$ for which there is an $a > 0$ with $X(t-\epsilon) \notin A$ and $X(t+\epsilon) \in A$ for all $0 < \epsilon < a$, or *exit times*, which have the reverse definition.

Let $\{T_n^{\to A}\}_{n\in\mathbb{Z}}$ denote the (non-decreasing) sequence of successive entrance times into A and $\{T_n^{A\to}\}_{n\in\mathbb{Z}}$ the exit times out of A, with the same numbering convention used for ordinary point processes. Let $N^{\to A}$ and $N^{A\to}$ be the point processes whose points are these sequences of entrance and exit times respectively. It is straightforward to show that these point processes are θ_t-compatible and assign each singleton a mass of 0 or 1 (that is to say, they are *simple*).

If, in addition to the above assumption on A, it is also true that the point processes $N^{\to A}$ and $N^{A\to}$ have finite and positive intensities then A is called *regular*. If $A, F \in \mathcal{S}$ are disjoint regular sets then define

- $N^{(A\to)F}$ to be the point process whose points consist of the subsequence of the exits $\{T_n^{A\to}\}_{n\in\mathbb{Z}}$ out of A after which X hits F before A;
- $N^{(\to A)F}$ to be the point process whose points consist of the subsequence of the entrances $\{T_n^{\to A}\}$ into A after which, once X has left A, it hits F before A;
- $N^{F(\to A)}$ to be the point process whose points consist of the subsequence of the first entrances into A after X has left F.

Figure 2.2 shows a sample path with the point $T_1^{(A\to)F}$ indicated.

The notation λ^N will be used to denote the intensity $\mathbb{E}[N(0,1]]$ of a general point process N, and the notation $\lambda^{A(\to F)}$ will be used in place of the more cumbersome $\lambda^{N^{A(\to F)}}$ in the case that N is of the form $N^{A(\to F)}$, with similar notation for the other entrance and exit point processes defined above.

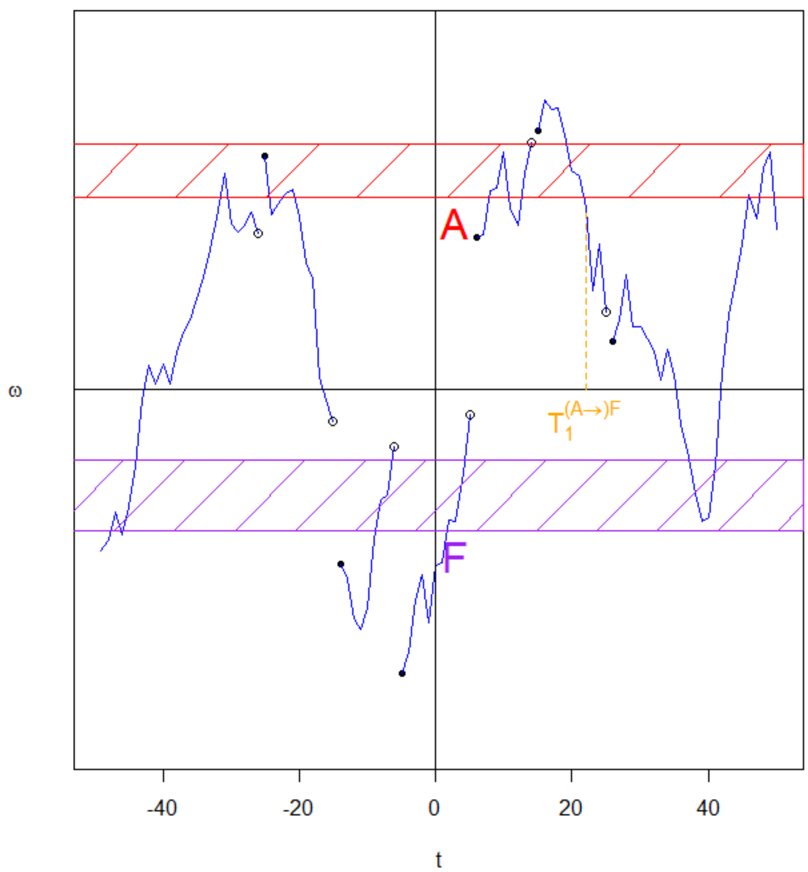

Figure 2.2: The point $T_1^{(A\to)F}$ for a sample path ω.

When considering more than one stationary point process at a time, such as $N^{F(\to A)}$ and $N^{F\to}$, it is useful to have a method of intelligibly switching between their associated Palm measures. The Neveu Exchange formula provides a means of doing so:

Theorem 2.1.1 *Let (N, θ_t, P) and (N', θ_t, P) be two stationary point processes with finite intensities. Then for any non-negative measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*

$$\mathbb{E}_0^N[f] = \frac{\mathbb{E}_0^{N'}[\int_{(0,T_1']}(f \circ \theta_t)N(dt)]}{\mathbb{E}_0^{N'}[N(0, T_1']]} \tag{2.1.9}$$

(see (1.3.6) of [3]).

2.2 The Markovian Setting

In this section assume that the process X is an ergodic Markov process with generator Q satisfying $\sum_{j \neq i} q_{ij} < \infty$ for each $i \in S$, stationary distribution μ, and define $\mathcal{F}_t := \sigma(X_u, u \leq t)$. Associate with X the *basic* point process N which counts the jump discontinuities of X and is given by (1.1.13) of [3]:

$$N(C) := \int_C 1_{S^2 \setminus \mathrm{diag}(S^2)}(X_{s^-}, X_s) m(ds), \quad C \in \mathbb{B}(\mathbb{R}) \tag{2.2.1}$$

where m is the counting measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. From this the *embedded discrete time chain* Y associated with X can be formed as $Y_n := X_{T_n} = X_0 \circ \theta_{T_n}$. This is the chain obtained by winnowing from X all information about its sojourn times.

Let $A, D \in \mathcal{S}$ be disjoint regular sets, let $\tau_n := T_n^{D(\to A)}$, and define the *watched process* W by $W_n := X_0 \circ \theta_{\tau_n}$. Observing W is equivalent to "watching" the process X only when it returns to A after having left D.

Recall the strong Markov property for stopping times τ:

$$P[\theta_\tau^{-1} G | \mathcal{F}_\tau] = P_{X_\tau}[G] \text{ a.s. on } \{\tau < \infty\} \tag{2.2.2}$$

for all $G \in \mathcal{F}^+ := \sigma(X_u, u \geq 0)$. (Here $\mathcal{F}_\tau$ is the σ-field

$$\mathcal{F}_\tau := \sigma(\{E \in \mathcal{F}; E \cap \{\tau \leq t\} \in \mathcal{F}_t\})$$

with respect to which both τ and X_τ are measurable).

Since X is cadlag, the points τ_n, $n \in \mathbb{Z}_+$ are stopping times with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{R}}$. Furthermore, $\tau_n \leq \tau_{n+1}$ holds for all n. Accordingly, the sigma fields $\mathcal{F}_{\tau_n}$ form a filtration for W.

Note that by (2.1.3) $W_{n+1} = X_0 \circ \theta_{T_{n+1}^{D(\to A)}} = X_0 \circ (\theta_{T_1^{D(\to A)}} \circ \theta_{T_n^{D(\to A)}}) = W_1 \circ \theta_{\tau_n}$ for each n. Then for $i_0, \ldots, i_n = i, j \in S$, and $E = \{W_n = i, W_{n-1} = i_{n-1}, \ldots, W_0 = i_0\}$, the strong Markov property implies

$$\begin{aligned} P[\{W_{n+1} = j\} \cap E] &= \mathbb{E}[I_E \mathbb{E}[I_{\{W_{n+1}=j\}} | \mathcal{F}_{\tau_n}]] \\ &= \mathbb{E}[I_E \mathbb{E}[I_{\theta_{\tau_n}^{-1}\{W_1=j\}} | \mathcal{F}_{\tau_n}]] \\ &= \mathbb{E}[I_E P_{W_n}[W_1 = j | \mathcal{F}_{\tau_n}]] \\ &= P[E] P_i[W_1 = j]. \end{aligned}$$

That is, W is a Markov chain.

Let K be the (embedded) kernel whose entries consist of the instantaneous transition probabilities of X:

$$K_{ij} := \frac{q_{ij}}{\sum_{\ell \neq i} q_{i\ell}}, \quad i, j \in S, i \neq j. \tag{2.2.3}$$

This kernel corresponds to the chain Y. From it the kernel corresponding to W can be determined: first define the n^{th} entrance time $\tau_B(n)$ into an arbitrary set B by $\tau_B(0) \equiv 0$ and

$$\tau_B(n) := N(0, T_n^{\to B}] = k \quad \Longleftrightarrow \quad \sum_{i=1}^{k} \chi\{Y_i \in B, Y_{i-1} \notin B\} = n, \tag{2.2.4}$$

and for $i, j \in S$, let ${}_BK_{ij}^{(0)} := \chi\{i = j\}$, ${}_BK_{ij}^{(1)} = K_{ij}$ and

$${}_BK_{ij}^{(n)} := \sum_{k \notin B} {}_BK_{ik}^{(n-1)} K_{kj} = P_i[Y_n = j, \tau_B(1) \geq n] \quad n \geq 2. \tag{2.2.5}$$

This is the probability that the chain Y, starting from i, moves to j via a path that avoids B in the $n-1$ intermediate steps. Now define

$$_BG_{ij} := \sum_{n=1}^{\infty} {_BK_{ij}^{(n)}} \tag{2.2.6}$$

for $i, j \in S$. This specifies the probability of a transition from i to j via a path (of any length) which avoids B. In the case $B = \emptyset$ note that $_BK^{(n)} = K^n$ and so $_BG_{ij} = \sum_{n=1}^{\infty} K_{ij}^n$ is just the potential series. Note that if K is recurrent then $_BG$ satisfies

$$\sum_{j\in B} {_BG_{ij}} = P_i[\tau_B(1) < \infty] = 1, \quad i \in S.$$

With this notation, the transition matrix for W can be described:

Lemma 2.2.1 *For any disjoint, regular sets $A, D \in \mathcal{S}$, the watched process $W_n = X_0 \circ \theta_{T_n^{D(\to A)}}$ is a Markov chain with transition kernel*

$$P[W_{n+1} = j | W_n = i] = \sum_{k\in D} ({_DG_{ik}})({_AG_{kj}}), \quad i, j \in A.$$

Proof: It is enough to observe that, a one step transition for W from $i \in A$ to $j \in A$ consists of a path for Y which must hit D for a first time and subsequently hit A for a first time. ∎

For each $i \in S$ $q_i := \sum_{j\neq i} q_{ij}$ gives the jump rate out of i, and so the rate at which X transitions from i, via a path that hits D and subsequently hits A first at $j \in A$, is

$$q_{ij}^{D(\to A)} := q_i \sum_{k\in D} ({_DG_{ik}})({_AG_{kj}}). \tag{2.2.7}$$

Therefore, by stationarity,

$$\int \int_t^{t+h} N^{D(\to A)}(ds)dP$$
$$= \mathbb{E}[N^{D(\to A)}(t, t+h)]$$

$$
\begin{aligned}
&= P[N^{D(\to A)}(t, t+h] = 1] + o(h) \\
&= \sum_{i \in S, j \in A} P[X_{t+h} = j, N^{D(\to A)}(t, t+h] = 1 | X_t = i] P[X_t = i] + o(h) \\
&= \sum_{i \in S, j \in A} q_{ij}^{D(\to A)} P[X_t = i] h + o(h) \\
&= \int \int_t^{t+h} \sum_{i \in S, j \in A} \mu(i) q_{ij}^{D(\to A)} ds dP + o(h)
\end{aligned}
$$

for $t \in \mathbb{R}$ and $h > 0$ small. It follows from this that

$$N^{D(\to A)}(ds) dP = \sum_{i \in S, j \in A} \mu(i) q_{ij}^{D(\to A)} ds dP.$$

Consequently

$$\lambda^{D(\to A)} = \mathbb{E}\left[\int_0^1 N^{D(\to A)}(ds)\right] = \sum_{i \in S, j \in A} \mu(i) q_{ij}^{D(\to A)}$$

and

$$P_0^{D(\to A)}[X_0 = j] = \frac{1}{\lambda^{D(\to A)}} \mathbb{E}\left[\int_0^1 I_{\{j\}}(X_s) N^{D(\to A)}(ds)\right] = \frac{\sum_{i \in S} \mu(i) q_{ij}^{D(\to A)}}{\sum_{i \in S, k \in A} \mu(i) q_{ik}^{D(\to A)}}$$

(compare Equation (1.4.8) of [3]). By (2.2.2) and (2.1.5)

$$
\begin{aligned}
P_0^{D(\to A)}[H] &= \frac{1}{\lambda^{D(\to A)}} \mathbb{E}\left[\sum_{n \in \mathbb{Z}_+} I_H(\theta_{\tau_n}) I_{(0,1]}(\tau_n)\right] \\
&= \frac{1}{\lambda^{D(\to A)}} \sum_{n \in \mathbb{Z}_+} \mathbb{E}\left(\mathbb{E}\left[I_{\theta_{\tau_n}^{-1} H} \mid \mathcal{F}_{\tau_n}\right] I_{(0,1]}(\tau_n)\right) \\
&= \frac{1}{\lambda^{D(\to A)}} \sum_{n \in \mathbb{Z}_+} \mathbb{E}\left(P_{X_{\tau_n}}[H] I_{(0,1]}(\tau_n)\right)
\end{aligned}
$$

for any $H \in \mathcal{F}^+$. So, by the mean value formula ((1.3.3) of [3]) with $g = I_j$,

$$
\begin{aligned}
P_0^{D(\to A)}[H] &= \frac{\sum_{j \in A} P_j[H] \mathbb{E}\left[\sum_{n \in \mathbb{Z}_+} \chi\{X_{\tau_n} = j\} I_{(0,1]}(\tau_n)\right]}{\mathbb{E}\left[\sum_{n \in \mathbb{Z}_+} I_{(0,1]}(\tau_n)\right]} \\
&= \sum_{j \in A} P_j[H] P_0^{D(\to A)}[X_0 = j]. \qquad (2.2.8)
\end{aligned}
$$

Therefore

Lemma 2.2.2 *Let X be an ergodic Markov chain with stationary distribution μ and generator $Q = (q_{ij})_{i,j\in S}$, let $H \in \mathcal{F}^+$, and let $A, D \in \mathcal{S}$ be disjoint regular sets. Then*

$$P_0^{D(\to A)}[H] = \mathbb{E}_0^{D(\to A)}[P_{X_{\tau_0}}[H]] = \frac{\sum_{i\in S, j\in A} \mu(i) q_{ij}^{D(\to A)} P_j[H]}{\sum_{i\in S, j\in A} \mu(i) q_{ij}^{D(\to A)}}$$

where $q_{ij}^{D(\to A)}$ is given by (2.2.7).

Proof: The first equality is just (2.2.8), and the second equality has been established by the foregoing arguments. ∎

Note that Lemma 2.2.2 can be leveraged for computations of the form $P_0^{D\to}[H]$ or $P_0^{\to A}[H]$, for example, by identifying the associated Palm measures as $P_0^{D(\to D^c)}$ or $P_0^{A^c(\to A)}$ respectively.

2.3 Reversals

The main theorem of Section 2.4 is a statement about how processes behave when the films of their trajectories are "watched" back to front. The implied process here is the *backwards* or *reversed process*, which will be denoted $\tilde{X}$. Some technical details will be worked out and some notation made clear.

For $\omega \in \Omega$ and $F \in \mathcal{F}$, define the *reversal* $\tilde{\omega} : \mathbb{R} \to \mathbb{R}$ of ω (respectively the *reversal* of F) via

$$\tilde{\omega}(t) := \lim_{s\uparrow -t} \omega(s), \quad \tilde{F} := \{\tilde{\omega}; \omega \in F\}.$$

Thus $\tilde{\omega}$ is the cadlag version of the map $\mathbb{R} \ni t \mapsto \omega(-t)$. Notice that $\omega \mapsto \tilde{\omega}$ defines a bijection of Ω. Define the *reversed process* $(\tilde{X}_t)_{t\in\mathbb{R}}$ by $\tilde{X}_t(\omega) = \tilde{\omega}(t)$.

The process $\tilde{X}$ is measurable with respect to $\mathcal{F}$. To see this, simply note that since Ω consists precisely of the cadlag trajectories from $\mathbb{R}$ to S, for any Borel set C

the set $\tilde{X}_t^{-1}(C)$ is of the form

$$\{\omega \in \Omega; \tilde{\omega}(t) \in C\} = \{\omega \in \Omega; \lim_{s \uparrow -t} \omega(s) \in C\} = \bigcup_{\epsilon \in (0,1) \cap \mathbb{Q}} \bigcap_{s \in (-t-\epsilon, -t) \cap \mathbb{Q}} \{\omega \in \Omega; \omega(s) \in C\}$$

and hence an element of $\mathcal{F}$. Therefore all cylinder sets corresponding to the process $\tilde{X}$ lie within $\mathcal{F}$ and $\tilde{X}$ is measurable. This also verifies that $\tilde{F} \in \mathcal{F}$ for any $F \in \mathcal{F}$.

Definition 2.3.1 *Let $N = \sum_{n \in \mathbb{Z}} \delta_{T_n}$ be a point process on $(\Omega, \mathcal{F}, P)$. Then the time reversal of N is the point process $\tilde{N}$ on $(\Omega, \mathcal{F}, P)$ defined by $\tilde{N} = \sum_{n \in \mathbb{Z}} \delta_{-T_n}$.*

Note that the point process $\tilde{N}$ is not compatible with the flow $\{\theta_t\}$ but rather the flow $\tilde{\theta}_t := \theta_{-t}$; this is because, for $t \geq 0$ and $C \in \mathcal{B}(\mathbb{R})$, (2.1.2) implies

$$\begin{aligned}
\tilde{N} \circ \theta_t(C) &= \sum_{n \in \mathbb{Z}} \delta_{-T_n \circ \theta_t}(C) \\
&= \sum_{n \in \mathbb{Z}} \delta_{-T_{n+N(0,t]}+t}(C) \\
&= \sum_{k \in \mathbb{Z}} \delta_{-T_k}(C - t) \\
&= \tilde{N}(C - t)
\end{aligned}$$

with a similar result holding for $t \leq 0$ (see (1.1.4) of [3]). Of course, P is compatible with $\tilde{\theta}_t$, and hence $(\tilde{N}, \tilde{\theta}_t, P)$ is a stationary point process for every stationary point process (N, θ, P). It should also be clear that the point process corresponding to $\tilde{X}$ is $\tilde{N}$. The Palm probability corresponding to $(\tilde{N}, \tilde{\theta}_t, P)$ is denoted $\tilde{P}_0^N$, and the intensity $\mathbb{E}[\tilde{N}(0,1]]$ of the point process $\tilde{N}$ is denoted $\tilde{\lambda}$.

Lemma 2.3.2 *For all $\omega \in \Omega$, $t \in \mathbb{R}$ and all random variables $W : \Omega \to \mathbb{R}$*

a) $\widetilde{\theta_t^{-1}} = \tilde{\theta}_t^{-1}$;

b) $\widetilde{\theta_t \omega} = \tilde{\theta}_{-t} \tilde{\omega}, \quad \widetilde{\theta_W \omega} = \tilde{\theta}_{-W} \tilde{\omega}$;

c) $\widetilde{\theta_t F} = \tilde{\theta}_{-t} \tilde{F}, \quad \widetilde{\theta_W F} = \tilde{\theta}_{-W} \tilde{F} \quad F \in \mathcal{F}$.

Proof: Assertion a) is verified by the observation that

$$\widetilde{\theta_t^{-1}} \circ \tilde{\theta}_t = \widetilde{\theta_{-t}} \circ \tilde{\theta}_t = \theta_t \circ \theta_{-t} = \theta_0.$$

Recall that the basic point process described by (2.2.1) counts the jump discontinuities of X, and that these jumps do not accumulate. Thus for any $u \in \mathbb{R}$ there is some $\epsilon > 0$ such that

$$\begin{aligned}
\tilde{\theta}_{-t}\tilde{\omega}(u) &= \theta_t \lim_{s\uparrow -u} \omega(s) \\
&= \begin{cases} \theta_t\omega(-u) & -u \text{ a continuity point of } \omega \\ \theta_t\omega(-u-\epsilon) & -u \text{ a point of discontinuity of } \omega \end{cases} \\
&= \begin{cases} \omega(t-u) & -u \text{ a continuity point of } \omega \\ \omega(t-u-\epsilon) & -u \text{ a point of discontinuity of } \omega \end{cases} \\
&= \begin{cases} \omega(t-u) & -u-t \text{ a continuity point of } \omega(\cdot+t) \\ \omega(t-u-\epsilon) & -u-t \text{ a point of discontinuity of } \omega(\cdot+t) \end{cases} \\
&= \lim_{s\uparrow -u} \omega(s+t) \\
&= \lim_{s\uparrow -u} \theta_t\omega(s) \\
&= \widetilde{\theta_t\omega}(u).
\end{aligned}$$

Now, notice that the rule $\tilde{\theta}_t = \theta_{-t}$ remains valid with t replaced by the random variable W (recall the definition $\theta_W(\omega) := \theta_{W(\omega)}(\omega)$ preceding (2.1.1)). Moreover, since the coordinate process X is compatible with $\{\theta_t\}_{t\in\mathbb{R}}$,

$$\theta_W\omega(s) = X_s(\theta_W\omega) = X_0(\theta_s \circ \theta_{W(\omega)}(\omega)) = X_0(\theta_{s+W(\omega)}(\omega)) = \omega(s + W(\omega))$$

and hence the above arguments go through with t replaced by W, verifying b). Assertion c) follows immediately from b). ∎

For $A \subseteq S$, note that $-T_1^{A\rightarrow} = \tilde{T}_{-1}^{\rightarrow A}$ on the set $\Omega_0^{A\rightarrow} := \{T_0^{A\rightarrow} = 0\}$. Applying

(2.1.4) and part c) of Lemma 2.3.2 with $W = T_1^{A\to}$ yields

$$\begin{aligned}\widetilde{\theta^{-1}_{T^{A\to}_{-1}} F} &= \widetilde{\theta_{T^{A\to}_1} F} = \tilde{\theta}_{-T^{A\to}_1} \tilde{F} = \tilde{\theta}_{\tilde{T}^{\to A}_{-1}} \tilde{F} \\ &= \tilde{\theta}^{-1}_{\tilde{T}^{\to A}_1} \tilde{F} \qquad \text{on} \quad \Omega_0^{A\to} = \{T_0^{A\to} = 0\}. \end{aligned} \tag{2.3.1}$$

Let $(\tilde{\mathcal{F}}_t)_{t\in\mathbb{R}}$ denote the filtration corresponding to $\tilde{X}$.

Lemma 2.3.3 *If X is a stationary Markov process with steady state μ then for any $F \in \mathcal{F}$, $P[F] = P[\tilde{F}]$.*

Proof: It is sufficient to check the claim for F belonging to a π-system generating $\mathcal{F}$ (this is because the operations of intersection, union and complement all commute with the reversal operator $\tilde{\cdot}$). Thus let F be a cylinder set of the form

$$F = \{X_{t_1} \in H_1, \ldots, X_{t_n} \in H_n\}$$

for some $n \in \mathbb{N}$, some real numbers $t_1 < \cdots < t_n$ and some sets $H_1, \ldots, H_n \in \mathcal{S}$. Let Q be the generator for X and define

$$\tilde{q}_{ij} := \frac{\mu(j) q_{ji}}{\mu(i)}, \quad \tilde{Q} := (\tilde{q}_{ij})_{i,j\in S}. \tag{2.3.2}$$

This is the generator for $\tilde{X}$ and has steady state μ. Let $(\mathbf{P}(t))_{t\in\mathbb{R}}$ and $(\tilde{\mathbf{P}}(t))_{t\in\mathbb{R}}$ denote the transition semigroups for X and $\tilde{X}$ respectively:

$$\mathbf{P}_{ij}(t) = P[X_t = j | X_0 = i] = (e^{Qt})_{ij}, \quad \tilde{\mathbf{P}}_{ij}(t) = P[\tilde{X}_t = j | \tilde{X}_0 = i] = (e^{\tilde{Q}t})_{ij}.$$

Notice that

$$\mathbf{P}_{ij}(t) = \frac{\mu(j)}{\mu(i)} \tilde{\mathbf{P}}_{ji}(t).$$

Then

$$\begin{aligned} P[F] &= \sum_{(i_1,\ldots,i_n)\in H_1\times\cdots\times H_n} P[X_{t_1} = i_1] \prod_{k=1}^{n-1} P[X_{t_{k+1}} = i_{k+1} | X_{t_k} = i_k] \\ &= \sum_{(i_1,\ldots,i_n)\in H_1\times\cdots\times H_n} \mu(i_1) \prod_{k=1}^{n-1} \mathbf{P}_{i_k i_{k+1}}(t_{k+1} - t_k) \end{aligned}$$

$$= \sum_{(i_1,\dots,i_n)\in H_1\times\cdots\times H_n} \mu(i_n) \prod_{k=1}^{n-1} \tilde{\mathbf{P}}_{i_{k+1}i_k}(t_{k+1}-t_k)$$
$$= P[\tilde{F}].$$

∎

Note that the steady state for K is the measure $p(i) := c\mu(i)\sum_{\ell\neq i} q_{i\ell}$ where $c = \left(\sum_{i\in S,\ell\neq i}\mu(i)q_{i\ell}\right)^{-1} = \left(-\sum_{i\in S}\mu(i)q_{ii}\right)^{-1}$ is a normalizing constant, and note that by (2.3.2), $\sum_{\ell\neq i} q_{i\ell} = -q_{ii} = -\tilde{q}_{ii} = \sum_{\ell\neq i}\tilde{q}_{i\ell}$. So defining $\tilde{K}_{ij} := \frac{p(j)K_{ji}}{p(i)}$, $i \neq j$, yields

$$\tilde{K}_{ij} = \frac{\mu(j)\sum_{\ell\neq j} q_{j\ell}}{\mu(i)\sum_{\ell\neq i} q_{i\ell}} \frac{q_{ji}}{\sum_{\ell\neq j} q_{j\ell}} = \frac{\tilde{q}_{ij}}{\sum_{\ell\neq i} q_{i\ell}} = \frac{\tilde{q}_{ij}}{\sum_{\ell\neq i}\tilde{q}_{i\ell}}. \tag{2.3.3}$$

This is the kernel corresponding to the embedded chain for $\tilde{X}$. Let ${}_B\tilde{K}^{(n)}$ be defined as in (2.2.5) with K_{ij} replaced by $\tilde{K}_{ij}$, let ${}_B\tilde{G} := \sum_{n=0}^{\infty} {}_B\tilde{K}^{(n)}$, let $\tilde{q}_i := \sum_{\ell\neq i}\tilde{q}_{i\ell}$, and let

$$\tilde{q}_{ij}^{D(\to A)} := \tilde{q}_i({}_D G_A G)_{ij}, \quad i \in S, j \in A. \tag{2.3.4}$$

2.4 The Folk Theorem

The Folk Theorem of this section is meant to make precise the following result which has been applied elsewhere (see [2]): the large deviation path from the origin to a rare event can be obtained by observing the time reversal from the rare event back to the origin. This will be formulated more rigorously with the notation of section (2.1).

Example 2.4.1 (Gambler's Coin) *Consider a simple random walk on* $\mathbb{N}$*: a gambler repeatedly flips a coin with probability of heads and tails given by* p *and* $q = 1-p$ *respectively,* $p < q$*. If heads comes up the gambler receives a dollar, and if tails comes up she loses a dollar if she has one to lose (i.e.* 0 *is a reflecting state). Let* Y_n *be the amount of money she has in hand after the* n^{th} *coin flip. Thus the transition kernel for the process* $\{Y_n\}_{n\geq 0}$ *is*

$$K = \begin{bmatrix} q & p & 0 & 0 & \\ q & 0 & p & 0 & \cdots \\ 0 & q & 0 & p & \\ 0 & 0 & q & 0 & \ddots \\ & \vdots & & \ddots & \ddots \end{bmatrix},$$

and the well known stationary distribution π *for this chain is given by* $\pi(n) = \left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^n$*. How does the gambler accrue a large fortune?*

If we are interested in the path to a large number ℓ *we might try simulating paths and extracting those which arrive at* ℓ *for analysis. However, if* ℓ *is quite large then this could be quite time consuming and inefficient since the vast majority of paths would need to be discarded. The revelation of the Folk theorem is that it is possible to start from the rare event and "watch the film in reverse" to see what rare excursions are likely to look like.*

The time reversal of K is the kernel $\tilde{K}$ given by

$$\tilde{K}_{ij} = \frac{\pi(j)}{\pi(i)} K_{ji} = \left(\frac{p}{q}\right)^{j-i} K_{ji}$$

and so K is reversible; that is, $\tilde{K} = K$. Thus when watching the gambler's fortune evolve in reverse as it begins at some large value ℓ and drifts back toward 0 one sees roughly q transitions to the left for every p transitions to the right. This means that the path which lead to the rare excursion contained roughly q transitions to the right for every p transitions to the left!

The surprising conclusion is that, the most likely paths which lead to the rare excursion are ones in which the coin behaved as if the probabilities of its sides were permuted (see Figure 2.3). Here, the simulated reversed paths $\tilde{Y}$ remained close in some sense (to be made precise in Section 2.4.2) to the line segment

$$\tilde{y}(s) = \ell + (p-q)s, \quad 0 \le s \le \frac{\ell}{q-p} \tag{2.4.1}$$

In the following discussion, assume that A and D_ℓ, $\ell \in \mathbb{N}$, are regular sets for $\{X_t\}_{t\in\mathbb{R}}$ in S satisfying $A \cap D_\ell = \emptyset$ and

$$\liminf_{\ell\to\infty} \tilde{P}_0^{D_\ell\to}[\tilde{T}_1^{A\to} < \tilde{T}_1^{D_\ell\to}] > 0. \tag{2.4.2}$$

Theorem 2.4.2 (Folk Theorem) *Let $\{F_\ell\}_{\ell\in\mathbb{N}}$ be a collection of measurable sets satisfying*

$$\lim_{\ell\to\infty} \tilde{P}_0^{D_\ell\to}[\tilde{F}_\ell] = 1. \tag{2.4.3}$$

Then

$$\lim_{\ell\to\infty} P_0^{(A\to)D_\ell}[\theta^{-1}_{T^{A\to}_{-1}} F_\ell] = 1.$$

Proof: Recall that, by (2.3.1), $\widetilde{\theta^{-1}_{T^{A\to}_{-1}} F_\ell} = \tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell$ under $P_0^{(A\to)D_\ell}$. Since

$$P_0^{(A\to)D_\ell}[U] = \tilde{P}_0^{D_\ell(\to A)}[\tilde{U}] \quad U \in \mathcal{F},$$

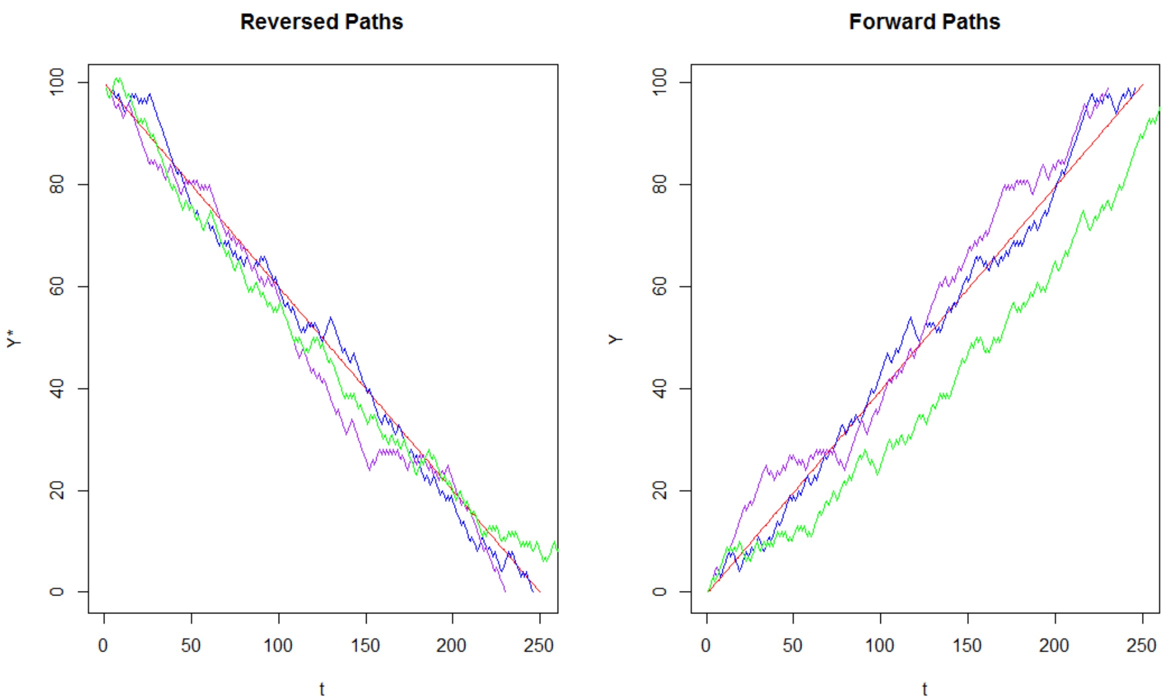

Figure 2.3: Reversal of Coin Flip Trajectories

it follows by the exchange formula (2.1.9) that

$$\begin{aligned} P_0^{(A\to)D_\ell}[\theta^{-1}_{T^{A\to}_{-1}} F_\ell] &= \tilde{P}_0^{D_\ell(\to A)}[\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell] \\ &= \tilde{\mathbb{E}}_0^{D_\ell(\to A)}[I_{\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell}] \\ &= \frac{\tilde{\mathbb{E}}_0^{D_\ell\to}\left[\int_{[0,\tilde{T}_1^{D_\ell\to})} I_{\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell} \circ \tilde{\theta}_t \tilde{N}^{D_\ell(\to A)}(dt)\right]}{\tilde{\mathbb{E}}_0^{D_\ell\to}[\tilde{N}^{D_\ell(\to A)}[0,\tilde{T}_1^{D_\ell\to})]}. \end{aligned}$$

Now, on the set $\{\tilde{T}_0^{D_\ell\to} = 0\}$, $\tilde{T}_1^{D_\ell(\to A)} = \tilde{T}_1^{\to A}$ and so the only point in the interval $[0, \tilde{T}_1^{D_\ell\to})$ to which $\tilde{N}^{D_\ell(\to A)}$ gives mass is $\tilde{T}_1^{\to A}$. Thus the above denominator is $\tilde{P}_0^{D_\ell\to}[\tilde{T}_1^{\to A} < \tilde{T}_1^{D_\ell\to}]$ and the numerator is

$$\tilde{\mathbb{E}}_0^{D_\ell\to}\left[\left(I_{\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell} \circ \tilde{\theta}_{\tilde{T}_1^{\to A}}\right) I_{\{\tilde{T}_1^{\to A} < \tilde{T}_1^{D_\ell\to}\}}\right] = \tilde{\mathbb{E}}_0^{D_\ell\to}\left[I_{\tilde{F}_\ell} I_{\{\tilde{T}_1^{\to A} < \tilde{T}_1^{D_\ell\to}\}}\right].$$

This means that

$$\begin{aligned} P_0^{(A\to)D_\ell}[\theta^{-1}_{T^{A\to}_{-1}} F_\ell] &= \frac{\tilde{P}_0^{D_\ell\to}[\tilde{F}_\ell \cap \{\tilde{T}_1^{A\to} < \tilde{T}_1^{D_\ell\to}\}]}{\tilde{P}_0^{D_\ell\to}[\tilde{T}_1^{A\to} < \tilde{T}_1^{D_\ell\to}]} \\ &\geq \frac{\tilde{P}_0^{D_\ell\to}[A_\ell]}{\tilde{P}_0^{D_\ell\to}[A_\ell] + \tilde{P}_0^{D_\ell\to}[\tilde{F}_\ell^c]} \end{aligned}$$

where $A_\ell = \tilde{F}_\ell \cap \{\tilde{T}_1^{A\to} < \tilde{T}_1^{D_\ell\to}\}$. Finally, using (2.4.3) and (2.4.2),

$$\liminf_{\ell\to\infty} P_0^{(A\to)D_\ell}[\theta^{-1}_{T^{A\to}_{-1}} F_\ell] \geq 1.$$

∎

To make intuitive sense of the Folk Theorem, treat the set D_ℓ as a rare event which gets rarer as $\ell \to \infty$ (perhaps it is a point in the state space of a Markov chain which is moving away), and treat A as the set on which the process X begins. Recall that the Palm measure $\tilde{P}_0^{D_\ell\to}$ concentrates all mass on the subset $\tilde{\Omega}_0^{D_\ell\to} = \{\tilde{T}_0^{D_\ell\to} = 0\}$ of Ω on which the backwards process $\tilde{X}$ has an exit from the rare set D_ℓ at time

0. Thus condition (2.4.3) is the requirement that the backwards process $\tilde{X}$ remains asymptotically within a tube $\tilde{F}_\ell$ of trajectories almost surely. So if one observes $\tilde{X}$ and then "runs the film in reverse", the expectation might be that the resulting forward trajectory will belong to the set of paths one gets by turning everything in $\tilde{F}_\ell$ around in time, i.e. F_ℓ.

This is inaccurate since turning everything in $\tilde{F}_\ell$ around in time would yield trajectories which arrive at the rare set D_ℓ at time 0. The idea is to start at the set A at time 0 then arrive at D_ℓ at time $T_1^{\to D_\ell}$, the first entrance time into D_ℓ. Thus each path $\omega \in F_\ell$ must be shifted by the amount of time it took the path $\tilde{\omega}$ to travel from the rare event to A, namely $\tilde{T}_1^{\to A} = -T_{-1}^{A\to}$. This is the reason that the conclusion of the Folk Theorem is expressed in terms of $\theta^{-1}_{T_{-1}^{A\to}} F_\ell$ rather than simply F_ℓ. Finally, the Palm measure $P_0^{(A\to)D_\ell}$ concentrates all its mass on paths which leave the starting set A at time 0 and subsequently reach the rare event before returning to A.

The Folk Theorem will be applied to the coin flipping example (Example 2.4.1) in Section 2.4.2. It was seen in that example that the reversed paths $\tilde{Y}$ remained close somehow to the line segment (2.4.1). It will be shown that the sets $\tilde{F}_\ell$ defined in 2.4.24, which are "tubes" of paths around a scaling of this line segment beginning at the rare point, satisfy (2.4.3). Trajectories of the reversed fortune process $\tilde{Y}$ sped up in time and scaled in space remain asymptotically within these tubes almost surely.

Thus the conclusion from the Folk Theorem is that the forward trajectories (properly scaled) remain asymptotically within a tube of trajectories which begin at the initial fortune and stay close to the same line segment.

2.4.1 Folk Theorem for Markov Chains

Working with Palm probabilities is useful for formalizing the key notions of the Folk Theorem, and for proving a general form of the result. However, the notation can be rather cumbersome and resultantly obscure many of the ideas. Since the main motivation is an application to discrete time Markov Chains, this Subsection will serve to provide computational devices for verifying some of the conditions associated with Theorem 2.4.2, as well as results more readily applied to the setting of interest.

In this Subsection suppose that X is an ergodic Markov process with generator Q and stationary measure μ. The following results allow condition (2.4.3) to be checked without appeal to Palm probabilities:

Lemma 2.4.3 *Let $\{\tilde{F}_\ell\}_{\ell\in\mathbb{N}}$ be a sequence of measurable sets from $\tilde{\mathcal{F}}^+$, and let $\tau := \tilde{T}_1^{D_\ell\rightarrow}$. Then*

$$\tilde{P}_0^{D_\ell\rightarrow}[\tilde{X} \in \tilde{F}_\ell] = \frac{\tilde{\lambda}}{\tilde{\lambda}^{D_\ell\rightarrow}} \sum_{i\in D_\ell} \sum_{j\in D_\ell^c} P_j[\tilde{F}_\ell]\mu(i)K_{ij} \tag{2.4.4}$$

$$= \frac{\sum_{i\in D_\ell}\sum_{j\in D_\ell^c} P_j[\tilde{F}_\ell]\mu(i)K_{ij}}{\sum_{i\in D_\ell}\sum_{j\in D_\ell^c}\mu(i)K_{ij}}. \tag{2.4.5}$$

Proof: Recall that with the notation of Section 2.3, $\tilde{\lambda}$ and $\tilde{\lambda}^{D_\ell\rightarrow}$ denote the intensities $\mathbb{E}[\tilde{N}(0,1]]$ and $\mathbb{E}[\tilde{N}^{D_\ell\rightarrow}(0,1]]$ of $\tilde{N}$ and $\tilde{N}^{D_\ell\rightarrow}$ respectively, and recall that K denotes the kernel associated with the embedded chain. Equation (1.5.3) of [3] implies, with $A = \tilde{F}_\ell = \{\tilde{X} \in \tilde{F}_\ell\}$ and $N = \tilde{N}^{D_\ell\rightarrow}$,

$$\tilde{P}_0^{D_\ell\rightarrow}[\tilde{X} \in \tilde{F}_\ell] = \lim_{t\to 0} P[\tilde{X}\circ\tilde{\theta}_\tau \in \tilde{F}_\ell | \tau \le t]. \tag{2.4.6}$$

Using the strong Markov property (2.2.2) write

$$\begin{aligned} P[\tilde{X}\circ\tilde{\theta}_\tau \in \tilde{F}_\ell, \tau \le t] &= \mathbb{E}(P[\tilde{X}\circ\tilde{\theta}_\tau \in \tilde{F}_\ell|\tilde{\mathcal{F}}_\tau]I_{\{\tau\le t\}}) \\ &= \mathbb{E}(P_{\tilde{X}_\tau}[\tilde{F}_\ell]I_{\{\tau\le t\}}) \\ &= \mathbb{E}\left(\sum_{i\in D_\ell}\sum_{j\in D_\ell^c} P_j[\tilde{F}_\ell]I_{\{\tilde{X}_\tau^- = i, \tilde{X}_\tau = j, \tau\le t\}}\right) \\ &= \sum_{i\in D_\ell}\sum_{j\in D_\ell^c} P_j[\tilde{F}_\ell]P[\tilde{X}_\tau^- = i, \tilde{X}_\tau = j, \tau \le t], \end{aligned}$$

and note that by (2.1.7)

$$P[\tilde{X}_\tau^- = i, \tilde{X}_\tau = j, \tau \le t] = P[\tilde{X}_\tau^- = i, \tilde{X}_\tau = j, \tau \le t, \tilde{N}(0,t] = 1] + o(t)$$

$$= P[\tilde{X}_{\tilde{T}_0} = i, \tilde{X}_{\tilde{T}_1} = j, \tilde{T}_1 \le t] + o(t), \quad i \in D_\ell, j \in D_\ell^c.$$

Moreover, by (1.4.27) of [3],

$$\begin{aligned}
&P[\tilde{X}_{\tilde{T}_0} = i, \tilde{X}_{\tilde{T}_1} = j, \tilde{T}_1 \le t] \\
&= P[\tilde{X}_{\tilde{T}_0} = i, \tilde{X}_{\tilde{T}_1} = j] - P[\tilde{X}_{\tilde{T}_0} = i, \tilde{X}_{\tilde{T}_1} = j, \tilde{T}_1 > t] \\
&= \tilde{\lambda}\mu(i)K_{ij}\int_0^\infty (1 - G_{ij}(s))ds - \tilde{\lambda}\mu(i)K_{ij}\int_t^\infty (1 - G_{ij}(s))ds \\
&= \tilde{\lambda}\mu(i)K_{ij}\int_0^t (1 - G_{ij}(s))ds
\end{aligned}$$

where $\tilde{\lambda}$ is the intensity of $\tilde{X}$ and

$$G_{ij}(s) := P[\tilde{T}_{n+1} - \tilde{T}_n \le s | \tilde{X}_{\tilde{T}_{n+1}} = j, \tilde{X}_{\tilde{T}_n} = i], \quad n \in \mathbb{Z}.$$

Thus, using (2.1.8), (2.4.6) becomes

$$\begin{aligned}
\tilde{P}_0^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] &= \lim_{t \to 0} \sum_{i \in D_\ell} \sum_{j \in D_\ell^c} \frac{P_j[\tilde{F}_\ell]\tilde{\lambda}\mu(i)K_{ij}\int_0^t (1 - G_{ij}(s))ds}{\tilde{\lambda}^{D_\ell \to} t} \\
&= \frac{\tilde{\lambda}}{\tilde{\lambda}^{D_\ell \to}} \sum_{i \in D_\ell} \sum_{j \in D_\ell^c} P_j[\tilde{F}_\ell]\mu(i)K_{ij} \lim_{t \to 0} \frac{1}{t}\int_0^t (1 - G_{ij}(s))ds \\
&= \frac{\tilde{\lambda}}{\tilde{\lambda}^{D_\ell \to}} \sum_{i \in D_\ell} \sum_{j \in D_\ell^c} P_j[\tilde{F}_\ell]\mu(i)K_{ij}(1 - G_{ij}(0^+)) \\
&= \frac{\tilde{\lambda}}{\tilde{\lambda}^{D_\ell \to}} \sum_{i \in D_\ell} \sum_{j \in D_\ell^c} P_j[\tilde{F}_\ell]\mu(i)K_{ij},
\end{aligned}$$

verifying (2.4.4). In the case that $\tilde{F}_\ell = \tilde{\Omega}$, this equation says

$$\frac{\tilde{\lambda}^{D_\ell \to}}{\tilde{\lambda}} = \sum_{i \in D_\ell} \sum_{j \in D_\ell^c} \mu(i)K_{ij},$$

which verifies (2.4.5). ∎

Lemma 2.4.4 *Let* $\tau := \tilde{T}_1^{D_\ell \to}$. *If* $\tilde{F}_\ell$ *is such that*

$$\lim_{\ell \to \infty} P[\tilde{X} \circ \tilde{\theta}_{\tilde{T}_1} \in \tilde{F}_\ell | \tilde{T}_1 = \tau] = 1 \tag{2.4.7}$$

then

$$\lim_{\ell \to \infty} \tilde{P}_{\mathbf{0}}^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] = 1.$$

Proof:

Let

$$\Delta(\ell, t) := \{\tilde{X}_{\tilde{T}_0} \in D_\ell, \tilde{X}_{\tilde{T}_1} \in D_\ell^c, \tilde{T}_1 \leq t\} = \{\tau = \tilde{T}_1, \tau \leq t\}$$

and

$$\Gamma(\ell) := \{\tilde{X} \circ \tilde{\theta}_{\tilde{T}_1} \in \tilde{F}_\ell\}.$$

Using (2.1.8) and (2.1.7) implies

$$\begin{aligned}
\lim_{t \downarrow 0} P[\tilde{X} \circ \tilde{\theta}_\tau \in \tilde{F}_\ell \mid \tau \leq t] &= \lim_{t \downarrow 0} \frac{P[\tilde{X} \circ \tilde{\theta}_\tau \in \tilde{F}_\ell, \tau \leq t]}{P[\tau \leq t]} \\
&= \lim_{t \downarrow 0} \frac{P[\tilde{X} \circ \tilde{\theta}_\tau \in \tilde{F}_\ell, \tau \leq t]}{\tilde{\lambda}^{D_\ell \to} t} \\
&= \lim_{t \downarrow 0} \frac{P[\tilde{X} \circ \tilde{\theta}_{\tilde{T}_1} \in \tilde{F}_\ell, \tilde{X}_{\tilde{T}_0} \in D_\ell, \tilde{X}_{\tilde{T}_1} \in D_\ell^c, \tilde{T}_1 \leq t]}{\tilde{\lambda}^{D_\ell \to} t} \\
&= \lim_{t \downarrow 0} \frac{P[\Gamma(\ell) \cap \Delta(\ell, t)]}{\tilde{\lambda}^{D_\ell \to} t}.
\end{aligned}$$

Furthermore, Equation (1.5.3) of [3] implies, with $A = \tilde{F}_\ell = \{\tilde{X} \in \tilde{F}_\ell\}$ and $N = \tilde{N}^{D_\ell \to}$,

$$\tilde{P}_{\mathbf{0}}^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] = \lim_{t \downarrow 0} P[\tilde{X} \circ \tilde{\theta}_\tau \in \tilde{F}_\ell | \tau \leq t].$$

So the above says that

$$\tilde{P}_{\mathbf{0}}^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] = \lim_{t \downarrow 0} \frac{P[\Gamma(\ell) \cap \Delta(\ell, t)]}{\tilde{\lambda}^{D_\ell \to} t}. \tag{2.4.8}$$

Note that (2.1.8) also implies

$$\lim_{t \downarrow 0} \frac{\tilde{\lambda}^{D_\ell \to} t}{P[\Delta(\ell, t)]} = \lim_{t \downarrow 0} \frac{P[\tau \leq t]}{P[\Delta(\ell, t)]} \frac{\tilde{\lambda}^{D_\ell \to} t}{P[\tau \leq t]} = \lim_{t \downarrow 0} \frac{P[\tau \leq t, \tilde{T}_1 \leq t] + o(t)}{P[\Delta(\ell, t)]} = 1.$$

Let $\epsilon \in (0,1)$ be arbitrary. Since $f_\ell := \mathbb{E}[I_{\Gamma(\ell)}|\tau_1, \tilde{T}_1]$ is equal to $P[\Gamma(\ell)|\tau = \tilde{T}_1]$ on the set $\{\tau = \tilde{T}_1\}$, the hypothesis and the preceding arguments entail that there exists an $L \in \mathbb{N}$ large enough so that $\ell \geq L$ implies

$$1 - \epsilon < f_\ell I_{\{\tau = \tilde{T}_1\}} + I_{\{\tau \neq \tilde{T}_1\}} < 1 + \epsilon \tag{2.4.9}$$

uniformly in $\tilde{\omega}$ and an $s(\ell) > 0$ small enough so that $t \leq s(\ell)$ implies

$$1 - \epsilon < \frac{P[\Delta(\ell,t)]}{\tilde{\lambda}^{D_\ell \to t}} < 1 + \epsilon. \tag{2.4.10}$$

Then, for $\ell \geq L$ and $t \leq s(\ell)$, multiplying both sides of (2.4.9) by $\frac{1}{\tilde{\lambda}^{D_\ell \to t}} I_{\Delta(\ell,t)}$ and taking expectations yields

$$(1-\epsilon)\frac{P[\Delta(\ell,t)]}{\tilde{\lambda}^{D_\ell \to t}} < \frac{\mathbb{E}[f_\ell I_{\Delta(\ell,t)}] + P[\Delta(\ell,t) \cap \{\tau \neq \tilde{T}_1\}]}{\tilde{\lambda}^{D_\ell \to t}} < (1+\epsilon)\frac{P[\Delta(\ell,t)]}{\tilde{\lambda}^{D_\ell \to t}},$$

after which applying (2.4.10) gives

$$(1-\epsilon)^2 < \frac{\mathbb{E}[f_\ell I_{\Delta(\ell,t)}]}{\tilde{\lambda}^{D_\ell \to t}} < (1+\epsilon)^2$$

(note that $\Delta(\ell,t) \subseteq \{\tau = \tilde{T}_1\}$). Since $\epsilon \in (0,1)$ it follows that

$$\left| \frac{\mathbb{E}[f_\ell I_{\Delta(\ell,t)}]}{\tilde{\lambda}^{D_\ell \to t}} - 1 \right| < 3\epsilon$$

whenever $\ell \geq L$ and $t \leq s(\ell)$. Also, observe that $\mathbb{E}[f_\ell I_{\Delta(\ell,t)}] = P[\Gamma(\ell) \cap \Delta(\ell,t)]$.

According to Equation (2.4.8), $s(\ell) > 0$ can be chosen small enough so that $t \leq s(\ell)$ also implies

$$\left| \tilde{P}_0^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] - \frac{P[\Gamma(\ell) \cap \Delta(\ell,t)]}{\tilde{\lambda}^{D_\ell \to t}} \right| < \epsilon.$$

Then $\ell \geq L$ and $t \leq s(\ell)$ imply

$$|\tilde{P}_0^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] - 1| \leq \left| \tilde{P}_0^{D_\ell \to}[\tilde{X} \in \tilde{F}_\ell] - \frac{P[\Gamma(\ell) \cap \Delta(\ell,t)]}{\tilde{\lambda}^{D_\ell \to t}} \right| + \left| \frac{P[\Gamma(\ell) \cap \Delta(\ell,t)]}{\tilde{\lambda}^{D_\ell \to t}} - 1 \right|$$

$$< \epsilon + \left| \frac{\mathbb{E}[f_\ell I_{\Delta(\ell,t)}]}{\tilde{\lambda}^{D_\ell \to} t} - 1 \right|$$
$$< 4\epsilon.$$

The conclusion is that $\lim_{\ell\to\infty} \tilde{P}_0^{D_\ell\to}[\tilde{X} \in \tilde{F}_\ell] = 1$. ∎

Corollary 2.4.5 *Suppose that*

$$\liminf_{\ell\to\infty} \left(\frac{\sum_{i\in S} \mu(i)\tilde{q}_{ij}^{D_\ell(\to A)}}{\sum_{i\in S,k\in A} \mu(i)\tilde{q}_{ik}^{D_\ell(\to A)}} \right) > 0 \tag{2.4.11}$$

for j in some nonempty set $C \subseteq A$. If $\{F_\ell\}_{\ell\in\mathbb{N}}$ is a collection of measurable sets belonging to $\tilde{\mathcal{F}}^+ = \sigma(\tilde{X}_u, u \geq 0)$ and satisfying Equation (2.4.7) then

$$\lim_{\ell\to\infty} P_j[\theta^{-1}_{T^{A\to}_{-1}} F_\ell] = 1 \quad \forall j \in C.$$

Proof: By Lemma 2.4.4 $\tilde{P}_0^{D_\ell\to}[\tilde{F}_\ell]$ converges to 1 as $\ell \to \infty$. Now, for an arbitrary set A the random variables $T_n^{(A\to)D_\ell}$ are not stopping times, and so are not directly amenable to the analysis of Section 2.2. However, $P_0^{(A\to)D_\ell}[U] = \tilde{P}_0^{D_\ell(\to A)}[\tilde{U}]$ for $U \in \mathcal{F}$, and $\tilde{T}_n^{D_\ell(\to A)}$ *are* stopping times for each $n \in \mathbb{Z}$. So applying Lemma 2.2.2 with $H = \widetilde{\theta^{-1}_{T^{A\to}_{-1}} F_\ell}$, which by (2.3.1) is equal to $\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell \in \tilde{\mathcal{F}}^+$ under $\tilde{P}_0^{D_\ell(\to A)}$, and $D = D_\ell$ yields

$$P_0^{(A\to)D_\ell}[\theta^{-1}_{T^{A\to}_{-1}} F_\ell] = \tilde{P}_0^{D_\ell(\to A)}[\widetilde{\theta^{-1}_{T^{A\to}_{-1}} F_\ell}] = \frac{\sum_{i\in S,j\in A} \mu(i)\tilde{q}_{ij}^{D_\ell(\to A)} P_j[\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell]}{\sum_{i\in S,j\in A} \mu(i)\tilde{q}_{ij}^{D_\ell(\to A)}}.$$

By (2.3.1) and Lemma 2.3.3 $P_j[\tilde{\theta}^{-1}_{\tilde{T}_1^{\to A}} \tilde{F}_\ell] = P_j[\theta^{-1}_{T^{A\to}_{-1}} F_\ell]$, and so taking the limit as $\ell \to \infty$ and applying Theorem 2.4.2 gives

$$1 = \lim_{\ell\to\infty} \frac{\sum_{i\in S,j\in A} \mu(i)\tilde{q}_{ij}^{D_\ell(\to A)} P_j[\theta^{-1}_{T^{A\to}_{-1}} F_\ell]}{\sum_{i\in S,j\in A} \mu(i)\tilde{q}_{ij}^{D_\ell(\to A)}}.$$

For each ℓ, let ρ_ℓ be the measure on A given by

$$\rho_\ell(j) := \frac{\sum_{i\in S} \mu(i)\tilde{q}_{ij}^{D_\ell(\to A)}}{\sum_{i\in S,k\in A} \mu(i)\tilde{q}_{ik}^{D_\ell(\to A)}}, \quad j \in A.$$

According to (2.4.11), the support of the measure $\rho := \liminf_{\ell\to\infty} \rho_\ell$ is C, and so

$$1 = \sum_{j \in C} \rho(j) \liminf_{\ell\to\infty} P_j[\theta_{T_{-1}^{A\to}} F_\ell].$$

Since this is a convex combination of positive terms summing to 1, the result follows. ∎

The Folk Theorem has been applied to a few networks [2, 19, 22].

Theorem 2.4.6 (Large Deviation Folk Theorem) *Let Y be the discrete time embedded chain associated with X. Let $\{z_\ell\}_{\ell\in\mathbb{N}}$ be a sequence of states satisfying $\mu(z_0) > 0$ and $\lim_{\ell\to\infty} \mu(z_\ell) \to 0$, let $\{\tilde{F}_\ell\}_{\ell\in\mathbb{N}}$ be a collection of measurable sets satisfying*

$$\lim_{\ell\to\infty} \frac{\sum_{j\neq z_\ell} K_{z_\ell j} P_j[\tilde{F}_\ell]}{\sum_{j\neq z_\ell} K_{z_\ell j}} = 1 \tag{2.4.12}$$

and let $E_\ell := \{T_0^{z_0\to} = 0, T_1^{\to z_\ell} < T_1^{\to z_0}\}$. Then

$$\lim_{\ell\to\infty} P_{z_0}[Y \in C_\ell | E_\ell] = 1 \tag{2.4.13}$$

where $C_\ell := \theta^{-1}_{T_{-1}^{z_0\to}} F_\ell$.

Proof: Consider the point process N of jumps out of z_0, and let $U := \{Y \in E_\ell\}$ be the set of trajectories which depart from z_0 at time zero and subsequently arrive at z_ℓ before z_0. Equation (1.4.7) of [3] defines the thinned process N_U as follows:

$$N_U(\omega, C) := \int_C 1_U(\theta_t\omega) N(\omega, dt), \quad C \in \mathcal{B}(\mathbb{R}).$$

Notice that with $A := \{z_0\}$ and $D_\ell := \{z_\ell\}$, $N_U = N^{(A\to)D_\ell}$. Moreover, notice that $P_{z_0} = P_0^N$ (see Section 1.7 in [3]). Therefore applying Equation (1.4.9) of [3] yields

$$P_{z_0}[Y \in C_\ell | E_\ell] = P_0^{N_U}[Y \in C_\ell] = P_0^{(A\to)D_\ell}[Y \in C_\ell].$$

By Lemma 2.4.3 and the hypothesis $\lim_{\ell\to\infty} \tilde{P}_0^{D_\ell\to}[\tilde{F}_\ell] = 1$, and so by Theorem 2.4.2 $\lim_{\ell\to\infty} P_0^{(A\to)D_\ell}[Y \in C_\ell] = 1$. The result then follows. ∎

2.4.2 Fluid Limits

A generalization of a Theorem of Kurtz [18] is often instrumental in finding an appropriate set $\tilde{F}_\ell$ with which to apply Theorem 2.4.6. The central idea is that the more "distant" the point in the state space at which the backwards process begins its journey, the more tightly the trajectories back to the origin, scaled in space and time, adhere to a deterministic path called the *fluid limit*. The following exposition follows [9].

For $T > 0$ let $D([0,T];\mathbb{R}^n)$ be the space of càdlàg trajectories from $[0,T]$ to $\mathbb{R}^n$. Assume that the countable state space S for X satisfies $S \subseteq \mathbb{R}^n_+$ (typically $S = \mathbb{N}^n$) and that the transition rates and directions for X are constant on each of the sets $D^s := \{x \in \mathbb{R}^n_+ \mid x_i = 0 \ \forall i \in s\}$ for s an element of the power set $\overline{\{1,\ldots,n\}}$ of $\{1,\ldots,n\}$. Thus $D^\emptyset = \{x \in \mathbb{R}^n_+ \mid \min_{i=1,\ldots,n} x_i > 0\}$ is the *interior* and $D^{\{1\}}$ is the first *boundary* of $\mathbb{R}^n_+$.

The process X will be assumed to be "nearest neighbor", i.e. $V := \{-1,0,1\}^n$ is the set of possible jump directions out of any state. The jump rate out of a state x in direction v will be denoted $\lambda_v(x)$. Thus $\lambda_v(x) = \lambda_v(y)$ whenever $x, y \in D^s$ for any $s \in \overline{\{1,\ldots,n\}}$. Moreover, since X lies in S, $\lambda_v(x) = 0$ whenever $x \in S$ but $x + v \notin S$.

The generator Q for X is therefore given by

$$q_{x,x+v} := \lambda_v(x), \quad x \in S, v \in V.$$

Here the process $Z_k(t) := \frac{1}{k}X(kt)$ scaled in time and space will be compared to a deterministic limit. The scaled process has the generator Q_k defined by

$$q_k(x, x + v/k) := k\lambda_v(x), \quad x \in S/k, v \in V.$$

Thus Z_k makes smaller jumps than those of X, at the accelerated rates $k\lambda_v(x)$. It should be clear that the process Z_k is Markovian. For $x \in \mathbb{R}^n_+$ define

$$B(x) := \{i \in \{1,\ldots,n\} \mid x_i = 0\} = s \quad \Leftrightarrow x \in D^s, \quad s \in \overline{\{1,\ldots,n\}}.$$

For $s \in \overline{\{1,\ldots,n\}}$ let $x_s \in \{x \mid B(x) = s\} = D^s$ and for $\beta \in \mathbb{R}^n$ define

$$L(s,\beta) = \sup_{\alpha\in\mathbb{R}^n} \left\{ \alpha\cdot\beta - \sum_{v\in V} \lambda_v(x_s)[\exp(\alpha\cdot v) - 1] \right\}. \tag{2.4.14}$$

Notice that $L(s,\beta) \geq 0$ (take $\alpha = 0$ in the above expression). Now for $x \in \mathbb{R}^n_+$, $\beta \in \mathbb{R}^n$ define

$$l(x,\beta) := \inf \left\{ \sum_{s\in\overline{B(x)}} \rho_s L(s,\beta_s) \mid \sum_{s\in\overline{B(x)}} \rho_s\beta_s = \beta, \bigwedge_{s\in\overline{B(x)}} \rho_s \geq 0, \sum_{s\in\overline{B(x)}} \rho_s = 1 \right\} \tag{2.4.15}$$

and define the functional I_x on functions $\phi \in D([0,T];\mathbb{R}^n)$ by

$$I_x(\phi) = \begin{cases} \int_0^T l(\phi(s),\phi'(s))ds & \phi \text{ absolutely continuous, } \phi(0) = x \\ +\infty & \text{otherwise.} \end{cases} \tag{2.4.16}$$

The following result from [9] will provide tubes $\tilde{F}_\ell$ to which the Folk Theorem can be applied.

Theorem 2.4.7 (Theorem 1.1 (ii) in [9]) *There exists a compact set $C \subset \mathbb{R}^n_+$ such that for each closed set F in $D([0,T];\mathbb{R}^n)$,*

$$\limsup_{k\to\infty} \frac{1}{k} \ln P_x[Z_k \in F] \leq -\inf_{\phi\in F} I_x(\phi)$$

uniformly in $x \in C$.

By choosing an appropriate closed set $F \subset D([0,T];\mathbb{R}^n)$ in Theorem 2.4.7 a Folk tube can be constructed. All such sets considered here will center around an object called the *fluid limit* associated with Q. Let

$$\beta_\emptyset := \sum_{v\in V} \lambda_v(x)v, \quad x \in D^\emptyset, \qquad \beta_{\{i\}} := \sum_{v\in V} \lambda_v(x)v, \quad x \in D^{\{i\}}. \tag{2.4.17}$$

These vectors give the drifts in the interior and on the i^{th} boundary. The *interior* fluid limit is a deterministic function p_∞ defined as the solution to the differential system

$$\frac{d}{dt} p_\infty(t) = \beta_\emptyset, \tag{2.4.18}$$

and having initial point $p_\infty(0) \in D^\emptyset$. The definition of the boundary fluid limits depends on a stability condition. Fix $i \in \{1, \ldots, n\}$ and let $x \in D^\emptyset$ be a point in the interior. Let V_i^+ and V_i^- be the set of jump directions $v \in V$ with positive or negative components in the direction of the i^{th} axis respectively. Then if the stability condition

$$\sum_{v \in V_i^+} \lambda_v(x) - \sum_{v \in V_i^-} \lambda_v(x) < 0, \quad x \in D^\emptyset \tag{2.4.19}$$

holds, the function $\nu_i : \mathbb{N} \to \mathbb{R}$ given by

$$\nu_i(k) := \nu_i(0) \left(\frac{\sum_{v \in V_i^+} \lambda_v(x)}{\sum_{v \in V_i^-} \lambda_v(x)} \right)^k \tag{2.4.20}$$

is a well defined probability measure for the appropriate normalizing constant $\nu_i(0)$. Condition (2.4.19) is an assurance that the process does not wander away in the direction of the i^{th} axis. When this condition holds, the i^{th} *boundary* fluid limit is the deterministic function p_∞^i defined as the solution to the differential system

$$\frac{d}{dt} p_\infty^i(t) = \nu_i(0)\beta_{\{i\}} + (1 - \nu_i(0))\beta_\emptyset =: \beta_{\emptyset,\{i\}}, \tag{2.4.21}$$

with initial point $p_\infty^i(0) \in D^{\{i\}}$. Note that for $x \in D^s$, $s = \emptyset, \{i\}$,

$$L(s, \beta_s) = \sup_{\alpha \in \mathbb{R}^n} \left\{ \sum_{v \in V} \lambda_v(x)[\alpha \cdot v - \exp(\alpha \cdot v) + 1] \right\}.$$

As α ranges over $\mathbb{R}^n$, the numbers $u_v := \alpha \cdot v$ range over $\mathbb{R}$, and so the above supremand is a convex combination of points lying on the graph of $f(u) = 1 + u - e^u$. This function is non positive, and thus any convex combination of the points of f is non-positive. Since $L(s, \beta_s) \geq 0$ it follows that $L(s, \beta_s) = 0$. If $s = \emptyset$ then the only element of $\overline{B(x)}$ is $s = \emptyset$ and hence (2.4.15) becomes

$$l(x, \beta_\emptyset) = L(\emptyset, \beta_\emptyset) = 0, \quad x \in D^\emptyset. \tag{2.4.22}$$

If instead $s = \{i\}$ then since $\beta_{\emptyset,\{i\}} = \nu_i(0)\beta_{\{i\}} + (1 - \nu_i(0))\beta_\emptyset$, the only convex combination $\rho_\emptyset L(\emptyset, \beta_\emptyset) + \rho_{\{i\}} L(\{i\})$ appearing in (2.4.15) is the one for which $\rho_{\{i\}} =$

$\nu_i(0)$ and $\rho_\emptyset = 1 - \nu_i(0)$. Thus

$$l(x, \beta_{\emptyset,\{i\}}) = \nu_i(0)L(\{i\}, \beta_{\{i\}}) + (1 - \nu_i(0))L(\emptyset, \beta_\emptyset) = 0, \quad x \in D^{\{i\}}. \tag{2.4.23}$$

Since the function l is strictly convex in its second coordinate (see [9]) it follows that $\beta_\emptyset$ and $\beta_{\emptyset,\{1\}}$ are the unique vectors satisfying (2.4.22) and (2.4.23).

Example 2.4.8 (Gambler's Coin Continued) *Returning to Example 2.4.1, let $\tilde{X}$ be the continuous time Markov process whose instantaneous transition probabilities are given by $\tilde{K}$ and whose holding times have rate 1. Thus $\tilde{Y}$ is the embedded chain associated with $\tilde{X}$. The interior fluid limit $\tilde{p}_\infty$ associated with $\tilde{K}$ can be computed from (2.4.18):*

$$\frac{d}{dt}\tilde{p}_\infty(t) = p - q$$

has solution $\tilde{p}_\infty(t) = \tilde{p}_\infty(0) + (p-q)t$. Let $\tilde{p}_\infty(0) = 1$ be the initial condition and let $T = \frac{1}{q-p}$ be the time point at which the fluid limit reaches 0. Then, for fixed $\epsilon > 0$,

$$\tilde{F} := \left\{ \sup_{0 \le t \le T} \left| \tilde{Z}_\ell(t) - \tilde{p}_\infty(t) \right| \le \epsilon \right\} = \left\{ \sup_{0 \le t \le T} \left| \frac{\tilde{Y}_{\tilde{N}(0,\ell t]}}{\ell} - \tilde{p}_\infty(t) \right| \le \epsilon \right\} \tag{2.4.24}$$

satisfies

$$\inf_{\phi \in \tilde{F}^c} I_x(\phi) > 0$$

for any $x > 0$ (the argument is similar to the arguments of the forthcoming Example). Then by Theorem 2.4.7

$$\lim_{\ell \to \infty} P_x[\tilde{Z}_\ell \in \tilde{F}^c] = 1$$

for any x in a compact set C. Take $C = [1-\delta, 1+\delta]$ so that if $\tilde{F}_\ell := \{\tilde{X}(\ell t)/\ell \in \tilde{F}^c\}$ and $\tilde{z}_\ell = \ell$,

$$\lim_{\ell \to \infty} P_{z_\ell}[\tilde{F}_\ell] = 1.$$

If $s = s(t) = \ell t$ is an acceleration of the time parameter by the factor ℓ, the above says that as $\ell \to \infty$ the backwards path $\tilde{X}(s)$ remains asymptotically within a "tube" *of radius $\epsilon\ell$ around the line segment*

$$\ell\tilde{p}_\infty(t) = \ell + (p-q)s, \quad 0 \le s \le \frac{\ell}{q-p}.$$

This makes precise the reason for (2.4.1).

Condition (2.4.12) holds:

$$\lim_{\ell\to\infty} \frac{\sum_{j\neq z_\ell} K_{z_\ell j} P_j[\tilde{F}_\ell]}{\sum_{j\neq z_\ell} K_{z_\ell j}} = \lim_{\ell\to\infty} \frac{\lambda P_{\ell+1}[\tilde{F}_\ell] + \mu P_{\ell-1}[\tilde{F}_\ell]}{\lambda+\mu} = 1,$$

and so by Theorem 2.4.6

$$1 = \lim_{\ell\to\infty} P_{z_0}[C_\ell \mid E_\ell] = \lim_{\ell\to\infty} P_{z_0}[\theta^{-1}_{T^{z_0\to}_{-1}} F_\ell \mid T^{z_0\to}_0 = 0, T^{\to z_\ell}_1 < T^{\to z_0}_1].$$

2.4.3 An Accounting Network

Consider the following example taken from [8]: a small insurance firm has an accounting department consisting of a senior accountant and her assistant the junior accountant. Operations in the department proceed as follows: all new claims arrive on the desk of the junior accountant, and do so according to a Poisson process with rate λ. The junior accountant processes the claims at rate μ_2, which subsequently either require more complicated calculations (with probability r_{21}) and thus proceed to the senior accountant, or else leave the system (with probability $r_{20} := 1 - r_{21}$). The senior accountant processes claims at rate μ_1 at which point they return to the desk of the junior accountant.

Identify the senior accountant's desk as node 1 and the junior accountant's desk as node 2 in this 2-node network. Let a point (x, y) in the state space $S := \mathbb{N}^2$ denote the the number x of claims on the senior accountant's desk and the number y of claims on the junior accountant's desk (see Figure 2.4).

Without loss of generality assume $\lambda + \mu_1 + \mu_2 = 1$ so that the jump rates can be interpreted as transition probabilities. Let $X(t)$ denote the state of the system at time t; clearly X is a Markov process. Let Y denote the discrete time embedded

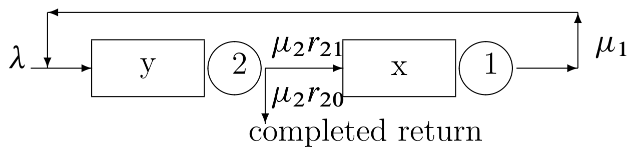

Figure 2.4: Flow Diagram for the Accountants' Network

Markov chain which describes the evolution of this network, and K its associated kernel. The transition diagram for the system is shown in Figure 2.5. The steady state of this Jackson network is

$$\mu(x, y) = (1 - \rho_1)(\rho_1)^x(1 - \rho_2)(\rho_2)^y \tag{2.4.25}$$

where

$$\rho_1 := \frac{\lambda_1}{\mu_1}, \rho_2 := \frac{\lambda_2}{\mu_2}, \lambda_1 := \frac{\lambda r_{21}}{r_{20}}, \lambda_2 := \frac{\lambda}{r_{20}},$$

and a simple calculation shows that $\frac{\mu(i)}{\mu(j)}K_{ij} = K_{ji}$; i.e. this process is reversible $(K = \tilde{K})$. Since the chain is reversible Figure 2.5 is the transition diagram for the backwards process as well. The assumptions $\lambda_1 < \mu_1$ and $\lambda_2 < \mu_2$ ensure stability of the network and will be retained.

It is of interest to study large deviations in the first coordinate, which correspond to large numbers of returns on the senior accountant's desk. This simplistic model with known stationary distribution μ facilitates the illustration of several points. However, the methodology is also applicable in situations where μ is unknown. Assume

$$\mu_1 + \lambda > \mu_2. \tag{2.4.26}$$

Then by stability

$$\mu_1 > \mu_2 - \lambda > \mu_2 - \mu_2 r_{20} = \mu_2 r_{21}; \tag{2.4.27}$$

i.e. the chain drifts northwest.

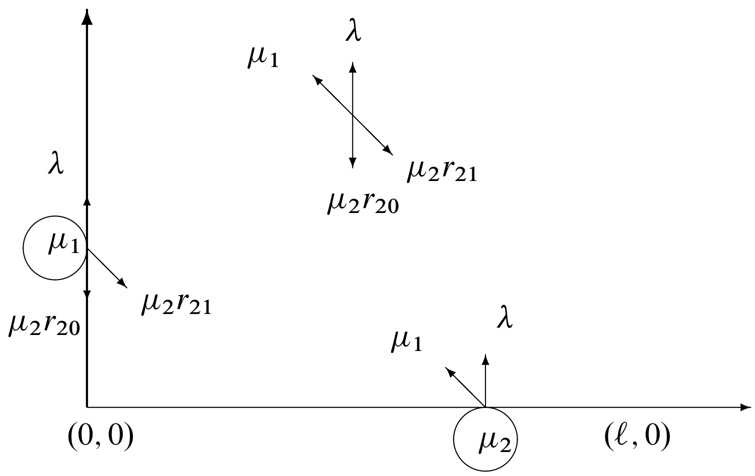

Figure 2.5: Transition diagram for the accountants' chain and its reversal

There are two boundaries for the process $\tilde{X}$ (the two axes $D^{\{1\}}$ and $D^{\{2\}}$). The interior fluid limit $\tilde{p}_\infty$ associated with $\tilde{K}$ satisfies (2.4.18):

$$\begin{aligned}\frac{d}{dt}\tilde{p}_\infty(t) &= \beta_\emptyset \\ &= \lambda(0,1) + \mu_2 r_{21}(1,-1) + \mu_2 r_{20}(0,-1) + \mu_1(-1,1) \\ &= (\mu_2 r_{20} - \mu_1, \lambda + \mu_1 - \mu_2),\end{aligned}$$

the solution of which is

$$\tilde{p}_\infty(t) = \tilde{p}_\infty(0) + t\beta_\emptyset.$$

Thus with the initial condition $\tilde{p}_\infty(0) = (1,0)$, $\tilde{p}_\infty$ is a line connecting the points

$$\tilde{p}_\infty(0) = (1,0) \quad \text{and} \quad \tilde{p}_\infty\left(\frac{1}{\mu_1 - \mu_2 r_{21}}\right) = \left(0, \frac{\lambda + \mu_1 - \mu_2}{\mu_1 - \mu_2 r_{21}}\right).$$

Let $T_1 := \frac{1}{\mu_1 - \mu_2 r_{21}}$; this is the time for the interior fluid limit to hit the y axis.

The steady state ν_1 for $\tilde{K}$ in the x direction far from the origin is given by (2.4.20):

$$\nu_1(k) = \left(1 - \frac{\mu_2 r_{21}}{\mu_1}\right)\left(\frac{\mu_2 r_{21}}{\mu_1}\right)^k \quad k \geq 0,$$

and so (2.4.21) becomes

$$\begin{aligned}\frac{d}{dt}\tilde{p}^1_\infty(t) &= \beta_{\emptyset,\{1\}} \\ &= \left(1-\frac{\mu_2 r_{21}}{\mu_1}\right)(\mu_2 r_{21}, \lambda-\mu_2) + \left(\frac{\mu_2 r_{21}}{\mu_1}\right)(\mu_2 r_{21}-\mu_1, \lambda+\mu_1-\mu_2) \\ &= (0, \lambda-\mu_2 r_{20})\end{aligned}$$

where $\tilde{p}^1_\infty$ is the boundary fluid limit along the y-axis. The solution to this system is

$$\tilde{p}^1_\infty(t) =: \tilde{p}^1_\infty(0) + t\beta_{\emptyset,\{1\}}.$$

Thus with the initial condition $\tilde{p}^1_\infty(0) = \tilde{p}_\infty(T_1)$, $\tilde{p}^1_\infty$ is a line connecting the points

$$\tilde{p}^1_\infty(0) = \left(0, \frac{\lambda+\mu_1-\mu_2}{\mu_1-\mu_2 r_{21}}\right) \quad \text{and} \quad \tilde{p}^1_\infty\left(\frac{\mu_2-\lambda-\mu_1}{(\mu_1-\mu_2 r_{21})(\lambda-\mu_2 r_{20})}\right) = (0,0).$$

Define $T = T_2 := \frac{1}{\mu_2 r_{20}-\lambda}$ and

$$\tilde{z}_\infty(t) := \begin{cases} \tilde{p}_\infty(t) & 0 \le t \le T_1 \\ \tilde{p}^1_\infty(t-T_1) & T_1 \le t \le T \end{cases}.$$

Thus $\tilde{z}_\infty(t)$ is a piecewise linear function connecting the points

$$\tilde{z}_\infty(0) = (1,0), \quad \tilde{z}_\infty(T_1) = \left(0, \frac{\lambda+\mu_1-\mu_2}{\mu_1-\mu_2 r_{21}}\right), \quad \text{and} \quad \tilde{z}_\infty(T) = (0,0).$$

It follows from (2.4.22) and (2.4.23) that $l(\tilde{z}_\infty(t), \tilde{z}'_\infty(t)) = 0$ for all $t > 0$. Therefore the functional (2.4.16) satisfies $I_x(\tilde{z}_\infty) = 0$ for the initial point $x = \tilde{z}_\infty(0)$. Define

$$\tilde{F}_\emptyset := \left\{\phi \in D([0,T];\mathbb{R}^2) \mid \sup_{t\in[0,\tau_\phi]} \|\phi(t) - (x+\beta_\emptyset t)\| > \epsilon\right\},$$

$$\tilde{F}_{\emptyset,\{1\}} := \left\{\phi \in D([0,T];\mathbb{R}^2) \mid \sup_{t\in[\tau_\phi,T]} \|\phi(t) - (\phi(\tau_\phi)+\beta_{\emptyset,\{1\}} t)\| > \epsilon\right\},$$

and

$$\tilde{F} := \tilde{F}_\emptyset \cup \tilde{F}_{\emptyset,\{1\}}$$

where

$$\tau_\phi := \inf\{t \geq 0 \mid \phi(t) \in D^{\{1\}}\}.$$

Then

$$\tilde{F}_\ell := \{\tilde{X}(\ell t)/\ell \in \tilde{F}^c\}$$

will be the "Folk tube" within which the scaled backwards process $\tilde{Z}_\ell$ is increasing likely to remain as $\ell \to \infty$. To demonstrate (2.4.12) Theorem 2.4.7 must be used, and it must be checked that

$$\inf_{\phi \in \tilde{F}} I_x(\phi) > 0. \tag{2.4.28}$$

There are some technical nuisances involved in verifying this condition: it is clear that $\tilde{z}_\infty \in \tilde{F}$, and that $\tilde{z}_\infty$ is the unique path satisfying $I_x(\tilde{z}_\infty) = 0$. What remains is to show that the value of $I_x(\phi)$ cannot be arbitrarily small for $\phi \in \tilde{F}^c$, and this involves handling many cases which needn't appear here.

Thus by Theorem 2.4.7 there is a compact set C such that, uniformly for $x \in C$,

$$\limsup_{\ell \to \infty} \frac{1}{\ell} \ln P_x[\tilde{Z}_\ell \in F] \leq -\inf_{\phi \in F} I_x(\phi).$$

This implies, by (2.4.28), that

$$\lim_{\ell \to \infty} P_x[\tilde{F}_\ell] = 1 \tag{2.4.29}$$

uniformly for x in a compact set containing z_ℓ.

This limit says that the time reversal starting from a distant point $(\ell, 0)$ will drift northwest remaining roughly within a tube of radius $\epsilon\ell$ around the line segment $\ell\tilde{z}_\infty(s/\ell), s \leq \ell T_1$ until it hits the y axis. It then travels down the y axis until it arrives at the point $(0, 0)$. See Figure 2.6 for a path simulated with the parameters $\lambda = 0.3, \mu_1 = 0.7, \mu_2 = 0.6, r_{21} = 0.2$ and $\ell = 150$.

Condition (2.4.12) holds:

$$\lim_{\ell \to \infty} \frac{\sum_{j \neq z_\ell} K_{z_\ell j} P_j[\tilde{F}_\ell]}{\sum_{j \neq z_\ell} K_{z_\ell j}} = \lim_{\ell \to \infty} \frac{\lambda P_{(\ell,1)}[\tilde{F}_\ell] + \mu_1 P_{(\ell-1,1)}[\tilde{F}_\ell]}{\lambda + \mu_1} = 1,$$

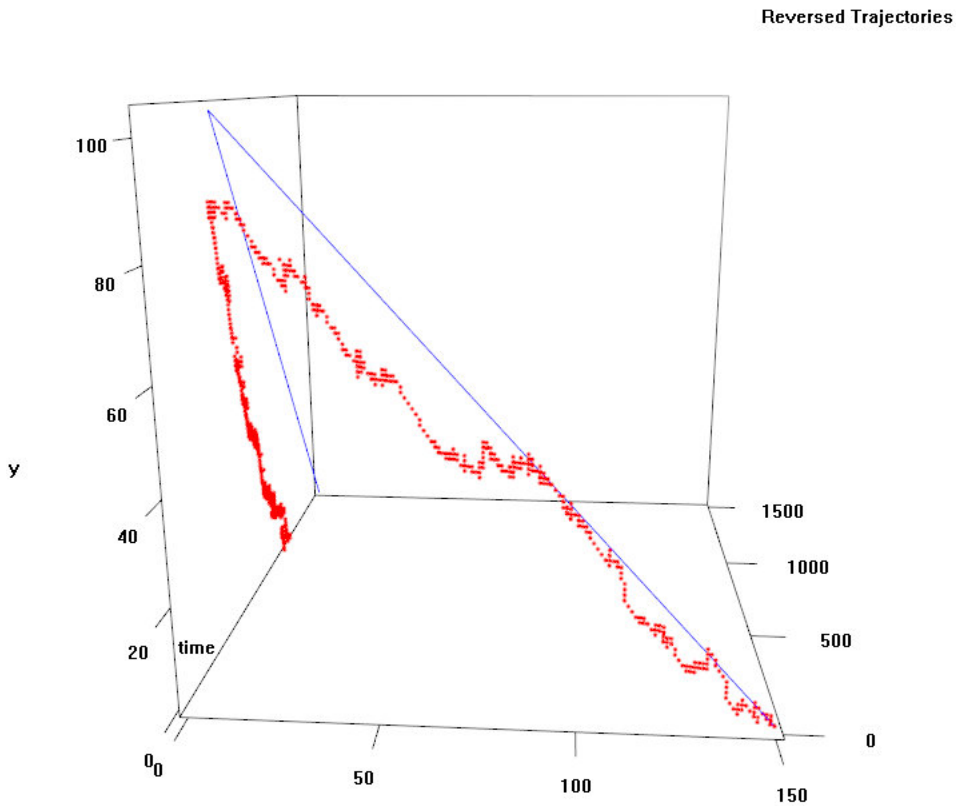

Figure 2.6: A simulated path $\tilde{Y}$ and $\tilde{p}_\infty$

and so by Theorem 2.4.6

$$\lim_{\ell\to\infty} P_{z_0}[Y \in C_\ell \mid T_0^{z_0\to} = 0, T_1^{\to z_\ell} < T_1^{\to z_0}] = 1.$$

Hence the large deviation path from $(0,0)$ has two segments, the first of which is a jitter up the y axis to a point at roughly $\left(0, \frac{\ell(\mu_1+\lambda-\mu_2)}{\mu_1-\mu_2 r_{21}}\right)$. To determine what the transitions of the forward process look like during this segment, reason as follows: the backwards process $\tilde{Y}$ begins this segment at roughly $\left(0, \frac{\ell(\mu_1+\lambda-\mu_2)}{\mu_1-\mu_2 r_{21}}\right)$, after which the x-component quickly reaches its local steady state ν. Thus the proportion of time spent by $\tilde{Y}$ on any of the rays $x = 0, x = 1, \ldots$ will approximately be given by $\nu(x)$, which then gives also the approximate proportion of time spent on these rays by the forward process Y, when near the y-axis. Therefore the proportion of transitions out of a point (x, y), in the north western direction say, is the product of the proportion of time spent on the vertical ray $x-1$ with the proportion of *south eastern* transitions experienced by $\tilde{Y}$ when at $(x-1, y+1)$:

$$\nu(x-1)\tilde{K}_{(x-1,y+1)(x,y)} = \nu(x-1)\mu_2 r_{21} = \mu_1 \nu(x).$$

Finally, the probability of that the forward process transitions from (x, y) to $(x-1, y+1)$ near $x = 0$ is given by dividing this proportion by the probability of being at the specified x-value, namely $\nu(x)$. This yields a probability μ_1 of a north western transition. The probability of other transitions are computed in a similar manner. The network near $x = 0$ seen by the accountants for the forward process leading to a large deviation is as in Figure 2.7. The accountants see new returns arriving at rate $\mu_2 r_{20}$ and returns exiting the system at rate λ. Thus the vertical drift for the large deviation path is $\mu_2 r_{20} - \lambda$; this is the negative of the drift associated with $K = \tilde{K}$. Compare this to Example 2.4.1 in which the path to the large fortune was one in which the coin behaved as if the sides were reversed.

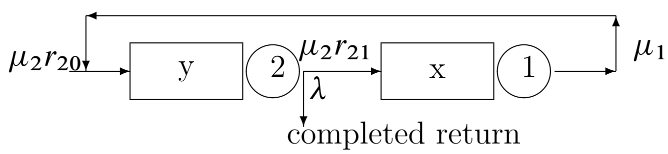

Figure 2.7: Transition rates for the large deviation path as it moves up the y axis

Chapter 3

Quasi-Stationary Measures for Sub-Stochastic Chains

3.1 The Yaglom Limit

It has been noted by several authors, perhaps most notably A.M. Yaglom, that many processes behave in an ergodic way in the short term but are eventually assured of experiencing a "death" of one kind or another. In many ways, it may be more realistic to analyze the limiting behaviour conditional on non-absorption rather than assuming that the governing probabilistic model remains valid indefinitely. In this section some definitions and notational conventions are laid out so that such analyses can proceed.

Let K be an irreducible, sub-stochastic kernel on the countable state space $S \cup \{0\}$ with *graveyard state* 0; that is, for each $i \in S$, $K_{i0} = 1 - \sum_{j \in S} K_{ij}$. Suppose that K is the kernel associated with the instantaneous transition probabilities of a Markov jump process $X = \{X_t\}_{t \in \mathbb{R}_+}$ with jump times $\{T_n\}_{n \in \mathbb{Z}}$. Let $Y = \{Y_n\}_{n \in \mathbb{N}}$ be defined by $Y_n := X_{T_n}$ where $\{T_n\}_{n \in \mathbb{Z}}$ are the points of X. Thus Y is a discrete time Markov chain on $S \cup \{0\}$ with transition kernel K. Assume that eventual absorption (or transition to 0) is certain: for all $i \in S$

$$P_i[\tau_0 < \infty] = 1 \quad \text{for} \quad \tau_0 := \inf\{n > 0; Y_n = 0\}. \tag{3.1.1}$$

Given that the chain started in state $i \in S$, let $\pi_n(\cdot|i)$ be the distribution of the process at time n conditional on non-absorption:

$$\pi_n(j) = \pi_n(j|i) := P_i[Y_n = j|Y_n \in S].$$

The more concise notation $\pi_n(j)$ will be preferred when i is understood, or unimportant.

Definition 3.1.1 *A Yaglom limit π for Y (or K) exists if, for any $i, j \in S$,*

$$\pi(j) = \lim_{n\to\infty} \pi_n(j|i). \tag{3.1.2}$$

Thus, when it exists, the Yaglom limit describes the behaviour of the non-absorbed process after long time periods and does not depend on the initial state i. There are interesting situations in which the limit above can exist *and* depend on i, but those cases are not treated here. Notice that by the series form of Scheffé's Theorem $\pi_n(A) \to \pi(A)$ for all $A \subseteq S$, and thus that $\pi_n \Rightarrow \pi$ (see the Corollary to Theorem 16.12 of [5]).

Lemma 3.1.2 *The Yaglom limit is a left eigenvector for K with eigenvalue*

$$\alpha := \lim_{n\to\infty} \frac{P_i[Y_n \in S]}{P_i[Y_{n-1} \in S]}, \ i \in S. \tag{3.1.3}$$

Proof: For any $i, j \in S$ and each n,

$$\begin{aligned}
\pi_n(j|i) &= \sum_{k\in S} \frac{P_i[Y_n = j, Y_{n-1} = k]}{P_i[Y_n \in S]} \\
&= \sum_{k\in S} \frac{P_i[Y_{n-1} = k]}{P_i[Y_n \in S]} K_{kj} \\
&= \frac{P_i[Y_{n-1} \in S]}{P_i[Y_n \in S]} \sum_{k\in S} \pi_{n-1}(k|i) K_{kj},
\end{aligned} \tag{3.1.4}$$

or

$$\frac{P_i[Y_n \in S]}{P_i[Y_{n-1} \in S]} = \frac{\sum_{k\in S} \pi_{n-1}(k|i)K_{kj}}{\pi_n(j|i)}.$$

Since $\pi_n \Rightarrow \pi$ taking the limit as $n \to \infty$ yields, by the Portmanteau lemma,

$$\lim_{n\to\infty} \frac{P_i[Y_n \in S]}{P_i[Y_{n-1} \in S]} = \frac{\sum_{k\in S} \pi(k)K_{kj}}{\pi(j)}.$$

Thus the limit in question exists, does not depend on j or i, and clearly is an eigenvalue of K corresponding to the left eigenvector π. ∎

It will be useful to invoke some of the terminology of [17] and [25]:

Definition 3.1.3 *A measure μ on S is called r-invariant for K if*

$$r\mu(j) = \sum_{i\in S} \mu(i)K_{ij} \tag{3.1.5}$$

holds for all j. A 1-invariant measure may simply be called invariant, and an α-invariant measure may be called quasi-invariant.

In the terminology of Definition (3.1.3) then, Lemma (3.1.2) says that any Yaglom limit is α-invariant for K with α defined by (3.1.3).

Corollary 3.1.4 *If Y has a Yaglom limit π then for all n*

$$\pi(i) = P_\pi[Y_n = i|Y_n \in S] \tag{3.1.6}$$

and

$$\alpha = \frac{P_\pi[Y_n \in S]}{P_\pi[Y_{n-1} \in S]} < 1. \tag{3.1.7}$$

Proof: This follows immediately from the fact that π is a left eigenvector for K:

$$P_\pi[Y_n = j] = \sum_{i\in S} \pi(i)K_{ij}^n = \alpha^n\pi(j) \tag{3.1.8}$$

and so $P_\pi[Y_n = j | Y_n \in S] = \frac{\alpha^n \pi(j)}{\sum_{k\in S} \alpha^n \pi(k)} = \pi(j)$ as well as

$$\frac{P_\pi[Y_n \in S]}{P_\pi[Y_{n-1} \in S]} = \frac{\alpha^n \sum_{j\in S} \pi(j)}{\alpha^{n-1} \sum_{j\in S} \pi(j)} = \alpha.$$

The fact that $\alpha < 1$ follows from summing (3.1.8) over all $j \in S$ and noting that the limit as $n \to \infty$ of the resultant sum is 0 by (3.1.1). ∎

There is another characterization of α which may be useful:

Lemma 3.1.5 *For any $i \in S$,*

$$\alpha = \lim_{n\to\infty} P_i[\tau_0 > n]^{1/n}.$$

Proof: Let $i \in S$ and $\epsilon > 0$ be arbitrary and let $L \in \mathbb{N}$ be large enough so that $n \geq L$ implies $\left|\alpha - \frac{P_i[Y_{n+1}\in S]}{P_i[Y_n \in S]}\right| < \epsilon$. Then for $n \geq L$

$$\begin{aligned} P_i[\tau_0 > n]^{1/n} &= \left(\frac{P_i[Y_n \in S]}{P_i[Y_{n-1} \in S]} \cdots \frac{P_i[Y_1 \in S]}{P_i[Y_0 \in S]} \right)^{1/n} \\ &\in ((\alpha - \epsilon)^{1-L/n} T(n), (\alpha + \epsilon)^{1-L/n} T(n)) \end{aligned}$$

where $T(n) = \left(\frac{P_i[Y_L \in S]}{P_i[Y_{L-1} \in S]} \cdots \frac{P_i[Y_1 \in S]}{P_i[Y_0 \in S]} \right)^{1/n}$ is a term going to 1 as $n \to \infty$. Thus taking the limit as $n \to \infty$ of the above expression shows that $\lim_{n\to\infty} P_i[\tau_0 > n]^{1/n} \in [\alpha - \epsilon, \alpha + \epsilon]$, and since ϵ was arbitrary the result follows. ∎

In this setting, one question is centrally important: will the process be absorbed before reaching a Yaglom limit? If so, the Yaglom limit is not of much practical use, and it is important to check this before making pronouncements about what to expect of the process after long time periods. The following well known result allows us to begin to answer the question.

Corollary 3.1.6 *If the Yaglom limit π exists and is the initial distribution for Y then τ_0 is geometrically distributed with parameter α:*

$$P_\pi[\tau_0 = n] = (1-\alpha)\alpha^n.$$

Proof: Summing (3.1.8) over all $j \in S$ gives $P_\pi[\tau_0 > n] = \alpha^n$, and the result follows by subtracting $P_\pi[\tau_0 > n+1] = \alpha^{n+1}$. ∎

Corollary (3.1.6) says that τ_0 is geometrically distributed when the chain is initially distributed according to the Yaglom limit, whereas Lemma (3.1.5) says that it is asymptotically so starting from any point $i \in S$.

The question of whether a Yaglom limit exists for a given kernel was addressed in [17] which gave sufficient conditions. These conditions, hereafter referred to as the *Kesten conditions*, are very much 1-dimensional restrictions. It is shown in [17] that when the Kesten conditions hold, when (3.1.1) holds, and when one step absorption is impossible from large enough states, an R^{-1}-invariant probability measure μ exists for K and satisfies

$$\mu(j) = \lim_{n\to\infty} \frac{K_{ij}^n}{\sum_{\ell \in S} K_{i\ell}^n} = \lim_{n\to\infty} \frac{P_i[Y_n = j]}{P_i[Y_n \in S]} = \lim_{n\to\infty} \pi_n(j|i) \tag{3.1.9}$$

for all $i, j \in S$ (take $i = k$ and $m = 0$ in equation (1.13) of that paper). Here R^{-1} is the value specified by (3.2.17). Thus, by (3.1.2) μ is a Yaglom limit for K with $R^{-1} = \alpha$.

Example 3.1.7 *Let $\{X_t\}_{t\in\mathbb{R}_+}$ be a birth-death process with jump rate $1 = \lambda + \mu$, $0 < \lambda < \mu$, and instantaneous transition probabilities*

$$p_{ij} = \begin{cases} \lambda & j = i+1 \\ \mu & j = i-1,\ i \geq 1. \end{cases}$$

This defines a sub-stochastic continuous time Markov chain with associated embedded *chain* $Y_n = X_{T_n}$, $n \geq 0$*; its kernel is*

$$K = \begin{bmatrix} 0 & \lambda & 0 & 0 & \cdots \\ \mu & 0 & \lambda & 0 & \cdots \\ 0 & \mu & 0 & \lambda & \\ & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{3.1.10}$$

Seneta and Vere-Jones discuss this kernel in their paper [24] and find the α*-invariant probability measure* $\pi(i) := (1 - \sqrt{\lambda/\mu})^2 i \left(\sqrt{\frac{\lambda}{\mu}}\right)^{i-1}$, $\alpha = 2\sqrt{\lambda\mu}$*, which does not satisfy the Kesten conditions (it is not aperiodic, for example) but nevertheless is a Yaglom limit for* K*. This kernel is central to some forthcoming examples.*

3.2 Deterministic Convergence to π

Equation (3.1.4) says that the conditional distribution π_n at time n satisfies the recursive relationship

$$\pi_n = C_{n-1}\pi_{n-1}K, \quad n \geq 1 \tag{3.2.1}$$

where $C_n := \frac{P[Y_n \in S]}{P[Y_{n+1} \in S]}$ is a quantity converging to α^{-1} as $n \to \infty$ (here π_n is treated as a row vector). This suggests an iterative, deterministic procedure for computing π_n and thus for approximating π, should it exist. To examine the rate of convergence of π_n to π, consider the *chi-square distance* introduced in [11] and given by

$$\chi_n^2 := \sum_{i \in S} \left(\frac{\pi_n(i)}{\pi(i)} - 1 \right)^2 \pi(i), \tag{3.2.2}$$

and the inner product structure imbued by π on the space of complex-valued functions over S:

$$\langle f, g \rangle_\pi := \sum_{i \in S} f(i)\overline{g(i)}\pi(i) \quad \text{for} \quad f, g : S \to \mathbb{R}. \tag{3.2.3}$$

The associated π-norm is defined by $||f||_\pi := \sqrt{\langle f, f \rangle_\pi}$. Then if $\Pi = diag(\pi)$ is the diagonal matrix whose entries are given by π and $f_n := \Pi^{-1}\pi_n^\top - \mathbf{1}$, the chi-square measure of the distance from the n^{th} iterate and π is just the square of the π-norm of f_n:

$$\chi_n^2 = ||f_n||_\pi^2. \tag{3.2.4}$$

Note that the adjoint of an operator A on the inner product space $\mathbb{F}(S) = (\mathbb{C}^S, \langle \cdot, \cdot \rangle_\pi)$, denoted by $\tilde{A}$, is the *reversal* with respect to π, i.e.

$$\tilde{A}_{ij} := \frac{\pi(j)}{\pi(i)}\overline{A_{ji}}. \tag{3.2.5}$$

Then $\tilde{A}$ and A^H are similar matrices since $\tilde{A} = \Pi^{-1}A^H\Pi$, and thus A and $\tilde{A}$ have the same eigenvalues (here A^H denotes the Hermitian of A).

The stochastic kernel K^π on S defined by $K_{ij}^\pi := K_{ij} + K_{i0}\pi(j)$ effectively resurrects dead chains and distributes them according to π, and satisfies

$$K^\pi = K + (\mathbf{1} - K\mathbf{1})\pi = KM + \mathbf{1}\pi \tag{3.2.6}$$

where

$$M := I - \mathbf{1}\pi. \tag{3.2.7}$$

Right multiplying both sides of (3.2.1) by $\mathbf{1}$ gives $C_{n-1}^{-1} = \pi_{n-1}K\mathbf{1}$ which implies $-C_{n-1}^{-1}\pi = \pi - \pi_{n-1}K\mathbf{1}\pi - \pi = \pi_{n-1}(\mathbf{1} - K\mathbf{1})\pi - \pi$ and so

$$\begin{aligned} f_n^\top &= \pi_n \Pi^{-1} - \mathbf{1}^\top \\ &= C_{n-1}(\pi_{n-1}K - C_{n-1}^{-1}\pi)\Pi^{-1} \\ &= C_{n-1}(\pi_{n-1}K + \pi_{n-1}(\mathbf{1} - K\mathbf{1})\pi - \pi)\Pi^{-1} \\ &= C_{n-1}(\pi_{n-1}K^\pi - \pi)\Pi^{-1} \\ &= C_{n-1}(\pi_{n-1}\Pi^{-1}(\tilde{K}^\pi)^\top - \mathbf{1}^\top) \end{aligned}$$

or, since $\mathbf{1}^\top = \mathbf{1}^\top(\tilde{K}^\pi)^\top$,

$$f_n = C_{n-1}\tilde{K}^\pi f_{n-1}. \tag{3.2.8}$$

Notice that, for any self adjoint matrix A,

$$\widetilde{K^n A} = \Pi^{-1}(K^n\overline{A})^\top \Pi = \Pi^{-1}\overline{A}^\top \Pi \Pi^{-1}(K^n)^\top \Pi = A\tilde{K}^n; \tag{3.2.9}$$

it is then immediate that

$$\tilde{K}^\pi = M\tilde{K} + \mathbf{1}\pi. \tag{3.2.10}$$

Lemma 3.2.1 *The matrix M is the projection onto the subspace of $\mathbb{F}(S)$ consisting of functions orthogonal to $\mathbf{1}$. Also,*

$$MKM = KM. \tag{3.2.11}$$

Proof: It is immediate that $\tilde{M} = M$ and $M^2 = M$. For any $g \in \mathbb{F}(S)$

$$\langle Mg, \mathbf{1}\rangle_\pi = \langle g - \mathbf{1}\pi g, \mathbf{1}\rangle_\pi = \langle g, \mathbf{1}\rangle_\pi - \langle \mathbf{1}\pi g, \mathbf{1}\rangle_\pi = 0,$$

and so M is the specified projection. Then $MKM = (K - \alpha\mathbf{1}\pi)M = KM$ verifies the specified equality. ∎

Notice that $KM\mathbf{1}\pi = 0$ and $\mathbf{1}\pi KM = \alpha\mathbf{1}\pi M = 0$ and so the cross terms that appear from expanding the powers $(K^\pi)^n = (KM + \mathbf{1}\pi)^n$ disappear. Also, $\mathbf{1}\pi$ is the projection onto $\mathbf{1}$ (the so called Perron projection in finite dimensions) and hence idempotent, and $(KM)^n = K^n M$ by (3.2.11). Therefore

$$(K^\pi)^n = K^n M + \mathbf{1}\pi, \qquad (\tilde{K}^\pi)^n = M\tilde{K}^n + \mathbf{1}\pi. \tag{3.2.12}$$

Lemma 3.2.2 *For each n, $f_n \perp \mathbf{1}$ and*

$$f_n = (C_{n-1}\cdots C_0)(\tilde{K}^\pi)^n f_0 = (C_{n-1}\cdots C_0)M\tilde{K}^n f_0.$$

Proof: Direct computation shows that

$$\langle f_n, \mathbf{1}\rangle_\pi = f_n^\top \Pi\mathbf{1} = (\pi_n\Pi^{-1} - \mathbf{1}^\top)\Pi\mathbf{1} = \pi_n\mathbf{1} - \pi\mathbf{1} = 0$$

since π_n and π are measures summing to 1. Then (3.2.8) and (3.2.12) imply

$$f_n = (C_{n-1}\cdots C_0)(\tilde{K}^\pi)^n f_0 = (C_{n-1}\cdots C_0)M\tilde{K}^n f_0.$$

∎

Suppose there exists a right eigenvector h for K with strictly positive entries corresponding to the eigenvalue α, represented as a column vector. Consider the assumption

$$\langle h, \mathbf{1}\rangle_\pi = \sum_{i\in S}\pi(i)h(i) = 1, \tag{3.2.13}$$

which, without loss of generality, is equivalent to the assumption $\sum_{i\in S}\pi(i)h(i) < \infty$.

Lemma 3.2.3 *a) The vector $\mathbf{1}$ is a right eigenvector for $\tilde{K}$ with eigenvalue α;*

b) If assumption (3.2.13) holds then the steady state ϕ for $\frac{1}{\alpha}\tilde{K}$ is given by

$$\phi(i) := \pi(i)h(i), \quad i \in S; \tag{3.2.14}$$

c) If K is aperiodic and assumption (3.2.13) holds then for each $i \in S$

$$h(i) = \lim_{n\to\infty} \frac{P_i[\tau_0 > n]}{\alpha^n}; \tag{3.2.15}$$

d) If $v \in \ker(K - \lambda I)^n$ for some $\lambda \neq \alpha$ then $v \in \ker(K^\pi - \lambda I)^n$ and $v \perp \mathbf{1}$.

e) If $v \in \ker(\tilde{K} - \lambda I)^n$ for some $\lambda \neq \alpha$ then $v \perp h$.

Proof: The calculation

$$\tilde{K}\mathbf{1} = \Pi^{-1}K^\top \Pi \mathbf{1} = \Pi^{-1}K^\top \pi^\top = \Pi^{-1}(\pi K)^\top = \Pi^{-1}(\alpha\pi)^\top = \alpha\mathbf{1}$$

verifies a). Under assumption (3.2.13) ϕ certainly defines a probability measure, and so

$$\phi \frac{1}{\alpha}\tilde{K} = \frac{1}{\alpha}\phi \Pi^{-1} K^\top \Pi = \frac{1}{\alpha} h^\top K^\top \Pi = \frac{1}{\alpha}\alpha h^\top \Pi = \phi$$

verifies b). If K is aperiodic then $\frac{1}{\alpha}\tilde{K}$ is (irreducible and) aperiodic and so

$$\left(\frac{1}{\alpha}\tilde{K}\right)^n_{ij} \to \phi(j), \quad i \in S$$

by part b). Now

$$P_i[\tau_0 > n] = \sum_{j \in S} K^n_{ij} = \frac{\alpha^n}{\pi(i)} \sum_{j\in S} \pi(j) \left(\frac{1}{\alpha}\tilde{K}\right)^n_{ji},$$

so by dominated convergence

$$\lim_{n\to\infty} \frac{P_i[\tau_0 > n]}{\alpha^n} = \frac{1}{\pi(i)} \sum_{j \in S} \pi(j)\phi(i) = h(i),$$

verifying c). Now if $v \in \ker(K - \lambda I)^n$ then

$$\begin{aligned} 0 &= \langle (K - \lambda I)^n v, \mathbf{1}\rangle_\pi \\ &= \sum_{i=0}^n \binom{n}{i} (-\lambda)^i \langle K^{n-i} v, \mathbf{1}\rangle_\pi \\ &= \sum_{i=0}^n \binom{n}{i} (-\lambda)^i \langle v, \tilde{K}^{n-i}\mathbf{1}\rangle_\pi \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=0}^{n} \binom{n}{i} (-\lambda)^i \alpha^{n-i} \langle v, \mathbf{1} \rangle_\pi \\
&= (\alpha - \lambda)^n \langle v, \mathbf{1} \rangle_\pi .
\end{aligned}
$$

If $\lambda \neq \alpha$ then $v \perp \mathbf{1}$ and by (3.2.12)

$$
\begin{aligned}
(K^\pi - \lambda I)^n v &= \sum_{i=0}^{n} \binom{n}{i} (-\lambda)^i (K^\pi)^{n-i} v \\
&= \sum_{i=0}^{n} \binom{n}{i} (-\lambda)^i K^{n-i} v \\
&= (K - \lambda I)^n v \\
&= 0
\end{aligned}
$$

so that d) holds. For $v \in \ker(\tilde{K} - \lambda I)^n$ the claim $\langle h, v \rangle_\pi = 0$ is verified by the calculation

$$
\begin{aligned}
0 &= \langle (\tilde{K} - \lambda I)^n v, h \rangle_\pi \\
&= \sum_{i=0}^{n} \binom{n}{i} (-\lambda)^i \langle \tilde{K}^{n-i} v, h \rangle_\pi \\
&= \sum_{i=0}^{n} \binom{n}{i} (-\lambda)^i \langle v, K^{n-i} h \rangle_\pi \\
&= \sum_{i=0}^{n} \binom{n}{i} (-\lambda)^i \alpha^{n-i} \langle v, h \rangle_\pi \\
&= (\alpha - \lambda)^n \langle v, h \rangle_\pi .
\end{aligned}
$$

∎

Note the distinction between part c) of Lemma 3.2.3 and Lemma 3.1.5; in particular, the former is a strictly stronger property.

Conditions for the existence, and uniqueness, of a positive h satisfying (3.2.13) are given in [25]. A sufficient condition is R-positivity of K, whose formulation requires a well known limit result: define the *potential series* associated with K to be

$$G_{ij}(z) := \sum_{n=0}^{\infty} K_{ij}^n z^n \tag{3.2.16}$$

where $K_{ij}^0 = \chi\{i = j\}$. Recall that K is irreducible. A routine application of Fekete's subadditivity lemma shows that

$$R^{-1} := \lim_{n\to\infty} (K_{ij}^n)^{1/n} \tag{3.2.17}$$

is a limit that exists and has the same value for all $i, j \in S$.

Lemma 3.2.4 *The radius of convergence of the potential series $G(z)$ is R.*

Proof: For $z \in \mathbb{C}$ let $y_n = (K_{ij}^n)^{1/n} z$. The root test says that $R^{-1}|z| = \lim_{n\to\infty} |y_n| < 1$ implies that $G_{ij}(z)$ is convergent and $R^{-1}|z| = \lim_{n\to\infty} |y_n| > 1$ implies that $G_{ij}(z)$ is divergent. The result follows. ∎

The kernel K is called R-*recurrent* if $G_{ij}(R) = \infty$, and R-*transient* otherwise. An R-recurrent matrix is called R-*positive* if none of the terms $K_{ij}^n R^n$ goes to zero in n, and R-*null* otherwise. These definitions have well known analogues in the stochastic case (where $R = 1$). To see the connection with Lemma 3.1.5, note that by part a) of Lemma 3.2.3

$$\begin{aligned}
(K_{ij}^n)^{1/n} &= \left(\frac{\pi(j)}{\pi(i)}\right)^{1/n} (\tilde{K}_{ji}^n)^{1/n} \\
&= \left(\frac{\pi(j)}{\pi(i)}\right)^{1/n} \left(\alpha \sum_k \tilde{K}_{jk}^{n-1} \left(\frac{1}{\alpha}\tilde{K}_{ki}\right)\right)^{1/n} \\
&\le \left(\frac{\pi(j)}{\pi(i)}\right)^{1/n} \left(\alpha \sum_k \tilde{K}_{jk}^{n-1}\right)^{1/n} \\
&= \left(\frac{\pi(j)}{\pi(i)}\right)^{1/n} \alpha.
\end{aligned}$$

Taking $n \to \infty$ in the above expression shows that $R^{-1} \le \alpha$. The reverse inequality may not hold in general, but Theorem D. of [25] says that if K is R-positive then

$\alpha = R^{-1}$ is the maximal eigenvalue of K, and that there exists a unique positive eigenvector h satisfying (3.2.13).

3.3 Conditions on π_0

Here conditions on π_0 sufficient to ensure that $\chi_n^2 \to 0$ will be given; this will demonstrate that the iterative approximation π_n of π converges in more than just the pointwise sense of (3.1.2).

Lemma 3.2.2 says that f_n is computed from f_0 by means of a large power of the matrix $\tilde{K}^\pi$. It is therefore reasonable to expect convergence to an element of the eigenspace of $\tilde{K}^\pi$, and since $f_n \perp \mathbf{1}$ for each n, a dominant eigenvalue of $\tilde{K}^\pi$ not equal to unity should describe the asymptotics of $\|f_n\|_\pi = \chi_n$. Retain the following assumption:

Assumption 3.3.1 *The initial distribution π_0 can be written as a linear combination of finitely many generalized (left) eigenvectors of K, among them π: for distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ of K ordered so that $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$, there exist non-zero scalars c, c_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, D_i$ such that*

$$\pi_0 = c\pi + \sum_{i=1}^{m} \sum_{j=1}^{D_i} c_{ij} w_{ij},$$

with $w_{ij}^\top \in \ker(K^\top - \lambda_i)^{n(i,j)}$ for some $n(i,j) \in \mathbb{N}$.

In a finite dimensional setting, D_i could be taken to be the algebraic multiplicity of the eigenvalue λ_i, and $w_{i1}, \ldots, w_{iD_i}$ the generalized left eigenvectors of K. In the infinite dimensional setting, there is no guaranty that an eigenspace $\ker(K^\top - \lambda I)$ is finite dimensional, but the assumption here is that finitely many generalized eigenvectors are sufficient to describe (via linear combination) the initial distribution π_0. Since

$$(\tilde{K} - \lambda_i I)^{n(i,j)} \Pi^{-1} w_{ij}^\top = \Pi^{-1} (K^\top - \lambda_i I)^{n(i,j)} w_{ij}^\top = 0,$$

i.e. $\Pi^{-1}w_{ij}^{\top} \in \ker(\tilde{K}-\lambda_i I)^{n(i,j)}$ for every i, j, it follows from (3.2.13) and part e) of Lemma 3.2.3 that

$$\pi_0 h = c\pi h + \sum_{i=1}^{m}\sum_{j=1}^{D_i} c_{ij} w_{ij} h = c + \sum_{i=1}^{m}\sum_{j=1}^{D_i} c_{ij} (\Pi^{-1}w_{ij}^{\top})^{\top}\Pi h = c,$$

and hence that

$$\pi_0 = (\pi_0 h)\pi + \sum_{i=1}^{m}\sum_{j=1}^{D_i} c_{ij} w_{ij}. \tag{3.3.1}$$

Thus

$$f_0 = \Pi^{-1}\pi_0^{\top} - \mathbf{1} = (\pi_0 h - 1)\mathbf{1} + \sum_{i=1}^{m}\sum_{j=1}^{D_i} c_{ij}\Pi^{-1}w_{ij}^{\top}.$$

Of course,

$$\pi_0 h = h^{\top}\pi_0^{\top} = h^{\top}\Pi(\Pi^{-1}\pi_0^{\top} - \mathbf{1}) + h^{\top}\Pi\mathbf{1} = \langle f_0, h\rangle_\pi + 1,$$

so

$$f_0 = \langle f_0, h\rangle_\pi \mathbf{1} + \sum_{i=1}^{m}\sum_{j=1}^{D_i} c_{ij}\Pi^{-1}w_{ij}^{\top}. \tag{3.3.2}$$

Now for $n \geq n(i) := \max_{j \leq D_i} n(i,j)$

$$\begin{aligned}
\tilde{K}^n \sum_{j=1}^{D_i} c_{ij}\Pi^{-1}w_{ij}^{\top} &= \sum_{j=1}^{D_i} c_{ij}(\tilde{K}-\lambda_i I + \lambda_i I)^n \Pi^{-1}w_{ij}^{\top} \\
&= \sum_{j=1}^{D_i} c_{ij}\sum_{k=0}^{n}\binom{n}{k}\lambda_i^{n-k}(\tilde{K}-\lambda_i I)^k\Pi^{-1}w_{ij}^{\top} \\
&= \lambda_i^n \sum_{j=1}^{D_i} c_{ij}\sum_{k=0}^{n(i,j)-1}\binom{n}{k}\lambda_i^{-k}(\tilde{K}-\lambda_i I)^k\Pi^{-1}w_{ij}^{\top} \\
&= \binom{n}{n(i)-1}\lambda_i^n g_{in}
\end{aligned}$$

for a vector $g_{in} \in \ker(\tilde{K}-\lambda_i I)^{n(i)}$ satisfying

$$g_{in} = \lambda_i^{-(n(i)-1)}(\tilde{K}-\lambda_i I)^{n(i)-1}\sum_{n(i,j)=n(i)} c_{ij}\Pi^{-1}w_{ij}^{\top} + O(n^{-1}) =: g_i + O(n^{-1}).$$

Note that $g_i \in \ker(\tilde{K} - \lambda_i I)$. Thus, for $n \geq \max_{i \leq m} n(i)$, (3.2.12) and $f_0 \perp \mathbf{1}$ implies

$$\begin{aligned}
(\tilde{K}^\pi)^n f_0 &= M \sum_{i=1}^m \binom{n}{n(i)-1} \lambda_i^n g_{in} \\
&= M \binom{n}{n(1)-1} \lambda_1^n \left(g_{1n} + \sum_{i=2}^m \binom{n}{n(1)-1}^{-1} \binom{n}{n(i)-1} \left(\frac{\lambda_i}{\lambda_1}\right)^n g_{in} \right) \\
&= \binom{n}{n(1)-1} \lambda_1^n (M g_1 + O(n^{-1})).
\end{aligned}$$

Note that if $n(1) = 1$ the $O(n^{-1})$ term can be replaced with $O\left(n^r \left|\frac{\lambda_2}{\lambda_1}\right|\right)$ for some r.

Now, $C_{n-1} \cdots C_0 = \frac{1}{P[\tau_0 > n]} = \frac{1}{\pi_0 K^n \mathbf{1}}$, so by Lemma 3.2.2

$$f_n = \frac{(\tilde{K}^\pi)^n f_0}{\pi_0 K^n \mathbf{1}} = \frac{\binom{n}{n(1)-1} \lambda_1^n (M g_1 + O(n^{-1}))}{\pi_0 K^n \mathbf{1}}. \tag{3.3.3}$$

Since (3.3.1) implies

$$\begin{aligned}
\pi_0 K^n \mathbf{1} &= \left((\pi_0 h)\pi + \sum_{i=1}^m \sum_{j=1}^{D_i} c_{ij} w_{ij} \right) K^n \mathbf{1} \\
&= (\pi_0 h)\alpha^n + \sum_{i=1}^m \sum_{j=1}^{D_i} c_{ij} ((K^\top)^n w_{ij}^\top)^\top \mathbf{1} \\
&= (\pi_0 h)\alpha^n + \sum_{i=1}^m \sum_{j=1}^{D_i} c_{ij} \lambda_i^n \sum_{k=0}^{n(i,j)-1} \binom{n}{k} \lambda_i^{-k} ((K^\top - \lambda_i)^k w_{ij}^\top)^\top \mathbf{1} \\
&= (\pi_0 h)\alpha^n + \sum_{i=1}^m \sum_{j=1}^{D_i} c_{ij} \lambda_i^n \sum_{k=0}^{n(i,j)-1} \binom{n}{k} \lambda_i^{-k} ((\tilde{K} - \lambda_i)^k \Pi^{-1} w_{ij}^\top)^\top \Pi \mathbf{1} \\
&= (\pi_0 h)\alpha^n + \sum_{i=1}^m \binom{n}{n(i)-1} \lambda_i^n g_{in}^\top \Pi \mathbf{1} \\
&= (\pi_0 h)\alpha^n + \binom{n}{n(1)-1} \lambda_1^n (g_1^\top + O(n^{-1})) \Pi \mathbf{1},
\end{aligned}$$

or

$$\frac{\pi_0 K^n \mathbf{1}}{\alpha^n} = \pi_0 h + \binom{n}{n(1)-1} \left(\frac{\lambda_1}{\alpha}\right)^n \langle g_1 + O(n^{-1}), \mathbf{1} \rangle_\pi, \tag{3.3.4}$$

it follows from (3.3.3) that

$$\left(\binom{n}{n(1)-1}\left(\frac{\lambda_1}{\alpha}\right)^n\right)^{-1} f_n = \frac{Mg_1 + O(n^{-1})}{\pi_0 K^n \mathbf{1}/\alpha^n} = \frac{Mg_1 + O(n^{-1})}{\pi_0 h + O\left(\binom{n}{n(1)-1}\right)} \longrightarrow \frac{Mg_1}{\pi_0 h}.$$

Theorem 3.3.2

Suppose that K is R-positive, and that the initial distribution π_0 can be represented as a linear combination of finitely many generalized left eigenvectors of K, as in (3.3.1), with $n(1)$ the largest order of the generalized eigenvectors in this linear combination corresponding to the dominant eigenvalue $|\lambda_1| < \alpha$. Then

a) there exists an eigenvector $g_1 \in \ker(\tilde{K} - \lambda_1 I)$ such that

$$\lim_{n\to\infty}\left(\binom{n}{n(1)-1}\left(\frac{\lambda_1}{\alpha}\right)^n\right)^{-1} f_n = \frac{Mg_1}{\pi_0 h}; \tag{3.3.5}$$

b) if $\|Mg_1\|_\pi < \infty$ then the asymptotic decay rate of χ_n to 0 is $\binom{n}{n(1)-1}\left(\frac{|\lambda_1|}{\alpha}\right)^n$.

Proof: Part a) has been argued above, and from (3.3.3) it follows that

$$\chi_n = \binom{n}{n(1)-1}\left(\frac{|\lambda_1|}{\alpha}\right)^n \frac{\|Mg_1\|_\pi + O(n^{-1})}{\pi_0 h + O\left(\binom{n}{n(1)-1}\right)},$$

which verifies b). ∎

Theorem 3.3.2 says that, properly scaled, f_n is approaching an eigenvector of $\tilde{K}^\pi$ with eigenvalue λ_1.

3.4 Comparison with Stochastic Case

In the fully stochastic case, an upper bound on the rate of convergence of the distribution π_n at time n to the steady state π is described in Section 2 of [11]. Theorem

3.3.2 expresses the subtlety of the approach needed in the sub-stochastic case. The recursive relationship specified by Lemma 3.2.2 implies

$$\chi_{n+1}^2 = C_n^2 \langle \tilde{K}^\pi f_n, \tilde{K}^\pi f_n \rangle_\pi = C_n^2 \left\langle \tilde{K}^\pi \frac{f_n}{||f_n||_\pi}, \tilde{K}^\pi \frac{f_n}{||f_n||_\pi} \right\rangle_\pi \chi_n^2,$$

and so applying the iterated inequality approach of [11] yields

$$\chi_{n+1}^2 \leq C_n^2 \beta_1(K^\pi \tilde{K}^\pi) \chi_n^2 \leq \ldots \leq (C_n \ldots C_1)^2 \beta_1(K^\pi \tilde{K}^\pi)^n \chi_0^2 \tag{3.4.1}$$

where

$$\beta_1(K^\pi \tilde{K}^\pi) := \sup\{\langle f, K^\pi \tilde{K}^\pi f \rangle_\pi ; ||f||_\pi = 1, f \neq \mathbf{1}\}.$$

The matrix $K^\pi \tilde{K}^\pi$ is the *multiplicative reversiblization* of K^π, and the minimax characterization of the eigenvalues of the real symmetric matrix $\Pi^{1/2} K^\pi \tilde{K}^\pi \Pi^{-1/2}$ implies that $\beta_1(K^\pi \tilde{K}^\pi)$ is the second largest eigenvalue of $K^\pi \tilde{K}^\pi$ (see 2.19 of [11]). Thus (3.4.1) gives an approximate upper bound of $\beta_1(K^\pi \tilde{K}^\pi)^{1/2}/\alpha$ on the decay rate for χ_n.

This bound can be unhelpful in the sub-stochastic case because it may happen that $\beta_1(K^\pi \tilde{K}^\pi)^{1/2}/\alpha > 1$, as is shown in the following example:

Example 3.4.1 *Let K be the 10×10 matrix given by*

$$K = \begin{bmatrix} 0 & \lambda & 0 & 0 & & 0 \\ \mu & 0 & \lambda & 0 & \cdots & 0 \\ 0 & \mu & 0 & \lambda & & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & 0 & \cdots & \mu & 0 & \lambda \\ 0 & 0 & \cdots & 0 & \mu & \lambda \end{bmatrix} \tag{3.4.2}$$

where $\lambda = 0.3$, $\mu = 0.7$, and let the initial distribution π_0 be uniform on $\{1, \ldots, 10\}$. Notice that 0 is an absorbing state. Using matlab to determine the eigenvalues of K yields $\alpha \approx 0.889$ and $\lambda_1 \approx 0.877$. The asymptotic rate given by Theorem 3.3.2

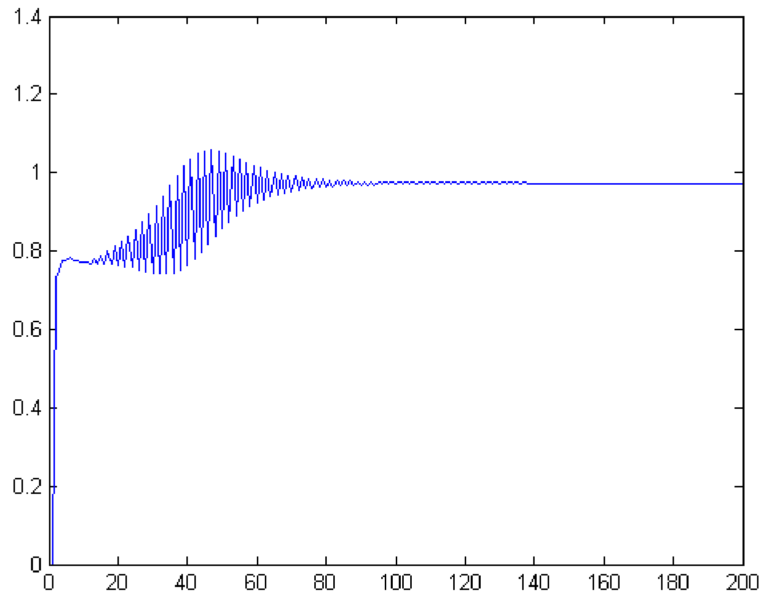

Figure 3.1: Ratio χ_{n+1}^2/χ_n^2 vs n.

is therefore approximately 0.987^n*, which was also verified by directly computing the values of* χ_{n+1}^2/χ_n^2 *for* $n = 1, ..., 200$ *(they converged to* $0.973 = (\lambda_1/\alpha)^2$*, see figure 3.1). However, the second largest eigenvalue of* $K^\pi \tilde{K}^\pi$ *is* $\beta_1(K^\pi \tilde{K}^\pi) \approx 0.856$ *and thus the asymptotic rate suggested by (3.4.1) is* ≈ 1.042*; while a valid upper bound, this inequality gives no information about whether* $\chi_n \to 0$*, which as a matter of fact does occur here.*

The reason that the methodology of [11] is inadequate to prove convergence is that the inequality (3.4.1) is formed "too early", and the resulting supremum $\beta_1(K^\pi \tilde{K}^\pi)$ ranges over too expansive a set. The effect of repeated applications of $\tilde{K}^\pi$ to f_0 cannot be ignored, and they are taken into account in Theorem 3.3.2 to achieve the crucial stabilization (3.3.5).

As mentioned in the discussion preceding Corollary 3.1.6, it is important to compare the rate at which substochastic chains are absorbed with the rate at which they converge to the quasi-stationary distribution so that one has a fair idea of what to expect at moderately large times n. Notice that when S is finite, Lemma 3.1.5 implies that $\lim_{n\to\infty} P_{\pi_0}[\tau_0 > n]^{1/n} = \alpha$ for any initial distribution π_0. Together with Theorem 3.3.2 this means that the convergence to quasi-stationarity is more rapid than absorption if and only if

$$|\lambda_1| < \alpha^2. \tag{3.4.3}$$

Example 3.4.2 *Let*

$$K = \frac{1}{10}\begin{bmatrix} 17/2 & 1/3 & 1/6 \\ 5/8 & 97/12 & 1/24 \\ 1/4 & 5/6 & 101/12 \end{bmatrix}.$$

Then $\pi = [1/2, 1/3, 1/6]$ *and* K *has the following Jordan decomposition* $K = VJV^{-1}$*:*

$$K = \begin{bmatrix} 2/3 & 0 & 1 \\ 1/2 & 1/20 & 1 \\ 1 & -1/10 & -5 \end{bmatrix}\begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 0.8 & 1 \\ 0 & 0 & 0.8 \end{bmatrix}\begin{bmatrix} 3/4 & 1/2 & 1/4 \\ -35/2 & 65/3 & 5/6 \\ 1/2 & -1/3 & -1/6 \end{bmatrix};$$

so the generalized eigenspace corresponding to $\lambda_1 = 0.8$ *is* 2*-dimensional. Thus the asymptotic rate of convergence of* χ_n *to zero is* $n\left(\frac{0.8}{0.9}\right)^n \approx n0.889^n$*. Here, the relation (3.4.3) holds and so one expects the chain to be distributed approximately according to* π *at moderately large values of* n*. In* 500 000 *simulations of a chain with kernel* K *and uniform initial distribution* π_0*,* 2850 *survived until time* 50 *(one expects roughly* $2578 \approx 500\ 000\alpha^{50}$*), and the empirical distribution of those that survived was approximately* $[0.4681, 0.3537, 0.1782]$ *which deviated from* π *(with respect to* $||\cdot||_\pi$*) by about* 0.06392 *(an asymptotic upper bound is* $0.1385 \approx 50(0.889)^{50}$*). This agrees with the theoretical results.*

Now

$$K^\pi = \begin{bmatrix} 9/10 & 3/45 & 3/90 \\ 1/8 & 17/20 & 1/40 \\ 1/20 & 1/10 & 17/20 \end{bmatrix}, \quad \tilde{K}^\pi = \begin{bmatrix} 9/10 & 1/12 & 1/60 \\ 1/10 & 17/20 & 1/20 \\ 1/10 & 1/20 & 17/20 \end{bmatrix}$$

and the second largest eigenvalue of $K^\pi \tilde{K}^\pi$ *is* $\beta_1(K^\pi \tilde{K}^\pi) \approx 0.681$*. So the rate implied by (3.4.1) is* $\beta_1(K^\pi \tilde{K}^\pi)^{1/2}/\alpha \approx 0.917$*. This overestimates the asymptotic rate, but is less egregious than Example 3.4.1 as the bound is sufficient to prove convergence. However, since* $\beta_1(K^\pi \tilde{K}^\pi)/\alpha > \alpha$*, using this bound would lead one to infer that absorption occurs faster than convergence to quasi-stationarity, which is not so.*

3.5 Continuous Case

The arguments of Section 3.2 essentially apply in the continuous case, though some technical aspects change. Consider a sub-stochastic continuous time Markov chain $\{X_t\}_{t\in\mathbb{R}}$, with infinitessimal generator $Q = \{q_{ij}\}_{i,j\in S}$ and transition semigroup $\mathbf{P} = (\mathbf{P}(t))_{t\geq\mathbf{0}}$; thus

$$\mathbf{P}_{ij}(t) = P[X_{s+t} = j \mid X_s = i], \quad i, j \in S, s, t \in \mathbb{R},$$

and the sub-stochasticity manifests in the condition

$$q_{i\mathbf{0}} := -\sum_{j\in S} q_{ij} \geq 0, \quad \forall i \in S \tag{3.5.1}$$

with strict inequality holding for at least one $i \in S$ (i.e. transition to the graveyard state $\mathbf{0}$ is possible from at least one state in S). As before, assume that eventual absorption is certain:

$$\lim_{t\to\infty} (\mathbf{P}(t)\mathbf{1})_i = \lim_{t\to\infty} P_i[X_t \in S] = 0, \quad \forall i \in S. \tag{3.5.2}$$

Recall that $\mathbf{P}(t) = \exp(Qt)$ is the solution to Kolmogorov's forward differential system

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{P}(t)Q. \tag{3.5.3}$$

As before let

$$\pi_t(j) = \pi_t(j|i) := P_i[X_t = j|X_t \in S] \tag{3.5.4}$$

with the notation $\pi_t(j)$ being preferred over $\pi_t(j|i)$ when mention of i is gratuitous. Suppose that the Yaglom limit π exists in the sense of definition (3.1.2) with the discrete index n replaced by the continuous index t:

$$\pi(j) = \lim_{t\to\infty} \pi_t(j \mid i), \quad i, j \in S. \tag{3.5.5}$$

Let $\pi_{\mathbf{0}}$ denote a row vector with 1 at entry i and 0's in all other entries, where i is any element of S. Then (3.5.4) says

$$\pi_t = \frac{\mathbf{P}_{i\cdot}^{\top}(t)}{\sum_{k\in S}\mathbf{P}_{ik}(t)} = \frac{\pi_{\mathbf{0}}\mathbf{P}(t)}{\pi_{\mathbf{0}}\mathbf{P}(t)\mathbf{1}}. \tag{3.5.6}$$

Lemma 3.5.1 *The Yaglom limit is a left eigenvector for* $\mathbf{P}(s)$ *with eigenvalue*

$$\alpha(s) := \lim_{t\to\infty} \frac{P_i[X_{s+t} \in S]}{P_i[X_t \in S]} = \lim_{t\to\infty} \frac{\pi_0 \mathbf{P}(t+s)\mathbf{1}}{\pi_0 \mathbf{P}(t)\mathbf{1}} \tag{3.5.7}$$

for all $s \geq 0$. *Moreover,* $\alpha(s) = \alpha^s$ *with* $\alpha := \alpha(1) \in (0,1)$, *and* π *is a left eigenvector for* Q *with eigenvalue* $\lambda = \ln \alpha$.

Proof: For $s, t \geq 0$, (3.5.6) implies

$$\begin{aligned} \pi_{t+s} &= \frac{\pi_0 \mathbf{P}(t+s)}{\pi_0 \mathbf{P}(t+s)\mathbf{1}} \\ &= \frac{\pi_0 \mathbf{P}(t)\mathbf{P}(s)}{\pi_0 \mathbf{P}(t)\mathbf{1}} \frac{\pi_0 \mathbf{P}(t)\mathbf{1}}{\pi_0 \mathbf{P}(t+s)\mathbf{1}} \\ &= \frac{\pi_0 \mathbf{P}(t)\mathbf{1}}{\pi_0 \mathbf{P}(t+s)\mathbf{1}} \pi_t \mathbf{P}(s), \end{aligned}$$

or

$$\frac{\pi_0 \mathbf{P}(t+s)\mathbf{1}}{\pi_0 \mathbf{P}(t)\mathbf{1}} = \frac{(\pi_t \mathbf{P}(s))_j}{\pi_{t+s}(j)}, \quad \forall j \in S.$$

For any sequence $\{t_n\}_{n\in\mathbb{N}}$ increasing to ∞, Scheffé's Theorem implies that $\pi_{t_n} \Rightarrow \pi$, and hence by the Portmanteau lemma the limit $t \to \infty$ of the above expression can be applied to conclude

$$\lim_{t\to\infty} \frac{\pi_0 \mathbf{P}(t+s)\mathbf{1}}{\pi_0 \mathbf{P}(t)\mathbf{1}} = \frac{(\pi \mathbf{P}(s))_j}{\pi(j)}, \quad \forall j \in S.$$

This verifies that $\alpha(s)$ is the specified eigenvalue. Let $\alpha := \alpha(1)$. Note $\alpha = \alpha \pi \mathbf{1} = \pi \mathbf{P} \mathbf{1} = P_\pi[X_1 \in S] > 0$. Then for every $n \in \mathbb{N}$

$$\alpha\pi = \pi \mathbf{P}(1) = \pi \mathbf{P}\left(\frac{1}{n} + \cdots + \frac{1}{n}\right) = \pi \mathbf{P}\left(\frac{1}{n}\right)^n = \alpha\,(1/n)^n\,\pi,$$

so $\alpha(1/n) = \alpha^{1/n}$. Then

$$\alpha^{m/n}\pi = \alpha(1/n)^m \pi = \pi \mathbf{P}\left(\frac{1}{n}\right)^m = \pi \mathbf{P}\left(\frac{m}{n}\right) = \alpha(m/n)\pi$$

so that $\alpha(m/n) = \alpha^{m/n}$ for all $m, n \in \mathbb{N}$. Since the semigroup $\mathbf{P}(s)$ is continuous in s it follows that $\alpha(s)$ is continuous and hence that

$$\alpha(s) = \alpha^s = e^{(\ln\alpha)s} =: e^{\lambda s}, \quad s \geq 0.$$

Thus

$$\alpha^t \pi = e^{\lambda t} \pi = \pi \mathbf{P}(t),$$

for all $t \geq 0$. Right multiplying by $\mathbf{1}$ and taking $t \to \infty$ in this expression demonstrates, in light of (3.5.2), that $\alpha < 1$, whereas applying (3.5.3) at $t = 0$ gives

$$\lambda \pi = \pi \mathbf{P}(0) Q = \pi Q.$$

∎

As in the discrete case, with

$$f_t := \Pi^{-1} \pi_t^\top - \mathbf{1},$$

the metric

$$\chi_t := \| f_t \|_\pi = \sqrt{\langle f_t, f_t \rangle_\pi} \tag{3.5.8}$$

gives the chi-square distance between the conditional distribution π_t and its limit π. A key ingredient to demonstrating $\chi_n \to 0$ and obtaining the corresponding decay rate for discrete chains was the recursive relationship (3.2.8) for f_n. No exact analogue exists in the continuous time case, but (3.5.3) gives rise to something similar. Differentiating (3.5.6) yields

$$\begin{aligned} \frac{d}{dt} \pi_t &= \frac{\pi_0 \mathbf{P}(t) Q \cdot \pi_0 \mathbf{P}(t) \mathbf{1} - \pi_0 \mathbf{P}(t) Q \mathbf{1} \cdot \pi_0 \mathbf{P}(t)}{(\pi_0 \mathbf{P}(t) \mathbf{1})^2} \\ &= \pi_t Q - \pi_t Q \mathbf{1} \cdot \pi_t \\ &= \pi_t Q (I - \mathbf{1} \pi_t). \end{aligned}$$

So if

$$M_t := I - \mathbf{1} \pi_t \tag{3.5.9}$$

this means that

$$\frac{d}{dt} \pi_t = \pi_t Q M_t = \pi_t (Q - \pi_t Q \mathbf{1} I). \tag{3.5.10}$$

Note

$$\pi_t Q \mathbf{1} = \frac{\pi_0 \mathbf{P}(t) Q \mathbf{1}}{\pi_0 \mathbf{P}(t) \mathbf{1}} = \frac{1}{\pi_0 \mathbf{P}(t) \mathbf{1}} \frac{d}{dt} \pi_0 \mathbf{P}(t) \mathbf{1} = \frac{d}{dt} \ln P[\tau > t]$$

where τ is the time to hit the cemetery state. Thus

$$\frac{d}{dt}\pi_t = \pi_t \left(Q - \frac{d}{dt} \ln P[\tau > t] I \right) \tag{3.5.11}$$

(the solution to this system of differential equations is

$$\pi_t = \pi_0 e^{Qt - \ln P[\tau > t] I} = \frac{\pi_0 e^{Qt}}{P[\tau > t]} = \frac{\pi_0 \mathbf{P}(t)}{\pi_0 \mathbf{P}(t) \mathbf{1}},$$

as required by (3.5.6)). Taking $t \to \infty$ in the second equality of (3.5.10) gives $\pi Q^\pi = \mathbf{0}^\top$ where Q^π is the generator given by

$$Q^\pi = Q - Q\mathbf{1}\pi = QM \tag{3.5.12}$$

with M given by (3.2.7). Therefore

$$\sum_{\ell \in \mathcal{S}} \pi(\ell)(q_{\ell j} + q_{\ell 0}\pi(j)) = 0 \quad \text{for every} \quad j \in S \tag{3.5.13}$$

(see (2.1) of [10]).

Lemma 3.5.2 *a) Both $\tilde{Q}\mathbf{1} = \lambda\mathbf{1}$ and $\tilde{Q}^\pi\mathbf{1} = \mathbf{0}$ hold;*

b) If $v \in \ker(Q - \beta I)^n$ for some $\beta \neq \lambda$ then $v \in \ker(Q^\pi - \beta I)^n$ and $v \perp \mathbf{1}$;

Proof: Assertion a) is verified by the computations

$$\tilde{Q}\mathbf{1} = \Pi^{-1} Q^\top \Pi \mathbf{1} = \Pi^{-1} Q^\top \pi^\top = \Pi^{-1}(\pi Q)^\top = \lambda \Pi^{-1} \pi^\top = \lambda \mathbf{1}$$

and

$$\tilde{Q}^\pi \mathbf{1} = \Pi^{-1}(Q^\pi)^\top \Pi \mathbf{1} = \Pi^{-1}(\pi Q^\pi)^\top = \mathbf{0}.$$

Assertion b) is proved exactly as in Lemma 3.2.3. ∎

Since

$$M f_t = f_t - \mathbf{1}\pi(\Pi^{-1}\pi_t^\top - \mathbf{1}) = f_t - \mathbf{1}\mathbf{1}^\top \pi_t^\top + \mathbf{1} = f_t,$$

it follows from (3.5.11) that

$$\begin{aligned}\frac{d}{dt} f_t &= \frac{d}{dt} M f_t \\ &= M\left(\Pi^{-1} \frac{d}{dt} \pi_t^\top - \mathbf{1}\right) \\ &= M\Pi^{-1}\left(Q^\top - \frac{d}{dt} \ln P[\tau > t] I\right)\pi_t^\top \\ &= M\left(\tilde{Q} - \frac{d}{dt} \ln P[\tau > t] I\right)\Pi^{-1}\pi_t^\top.\end{aligned}$$

As $M\tilde{Q}\mathbf{1} = \lambda M\mathbf{1} = \mathbf{0}$, the above says that

$$\frac{d}{dt} f_t = M\left(\tilde{Q} - \frac{d}{dt} \ln P[\tau > t] I\right) f_t = \left(\tilde{Q}^\pi - \frac{d}{dt} \ln P[\tau > t] I\right) f_t$$

and the solution to this system of differential equations is

$$f_t = M \frac{e^{\tilde{Q}t} f_0}{P[\tau > t]} = \frac{e^{\tilde{Q}^\pi t} f_0}{P[\tau > t]} \tag{3.5.14}$$

(compare Lemma 3.2.2). Thus

$$\chi_t = \|f_t\|_\pi = \frac{1}{P[\tau > t]} \|e^{\tilde{Q}t} f_0\|_\pi.$$

The rate at which $\chi_t \to 0$ can then be determined in a manner similar to the arguments leading to Theorem 3.3.2.

Chapter 4

Rare Events for Substochastic Chains

4.1 The Sustained Kernel

The Folk Theorem says that rare paths of a stationary process are described by the reversal of paths which begin at the rare destination. In practical terms, this suggests simulating the reversed process from the rare event and waiting until it reaches the set A of Theorem 2.4.2 (typically the origin). Trying to apply this to substochastic Markov chains leads to difficulties since, firstly, a substochastic chain can't very well be stationary, and secondly, there is some chance that the simulated chain (or its reversal) will be absorbed. Clearly, the Folk Theorem does not hold.

Nevertheless, the conviction that it is easier to view the chain backward in time starting from the rare event than to wait for the rare event to occur persists; indeed, in some respects this approach is all the more advisable in the substochastic setting since rarity is effectively compounded by the additional requirement to stay alive. To reconcile these observations one must employ the reversal of a modified kernel, and the fact that $\frac{1}{\alpha}\tilde{K}$ is a fully stochastic kernel (or that $\frac{1}{\alpha}\tilde{Q}$ a fully stochastic generator) is suggestive.

Let K be the embedded kernel of the substochastic Markov process $X = \{X_t\}_{t\in\mathbb{R}}$ on S, and let $Y = \{Y_n\}_{n\in\mathbb{Z}}$ be the corresponding embedded discrete time chain. Assume that the Yaglom limit π with corresponding eigenvalue α exists. The kernel which describes transition probabilities conditional upon absorption not occurring until some time n far into the future is

$$P_i[Y_1 = j \mid Y_n \in S] = \frac{P_i[Y_1 = j, Y_n \in S]}{P_i[Y_n \in S]} = K_{ij}\frac{P_j[Y_{n-1} \in S]}{P_i[Y_n \in S]}. \tag{4.1.1}$$

Clearly this is a stochastic kernel on S. Under assumption (3.2.13) taking the limit as $n \to \infty$ of (4.1.1) gives, by Lemma 3.2.3,

$$\lim_{n\to\infty} P_i[Y_1 = j \mid Y_n \in S] = \frac{h(j)}{\alpha h(i)}K_{ij} = K^h_{ij} \tag{4.1.2}$$

where h is the right eigenvector for K corresponding to α and satisfying $\langle h, \mathbf{1}\rangle_\pi = 1$ and

$$K^h_{ij} := \frac{K_{ij}h(j)}{\alpha h(i)} \tag{4.1.3}$$

is the *sustained kernel* of the chain. If it is known that a rare event will occur at some point far into the future, so that in particular the chain survives for a long time, the transition probabilities are described by K^h. It is therefore a chain with this kernel to which the Folk Theorem should be applied. Note that if one forms the measure on S given by $S \supseteq A \mapsto \sum_{j\in A} K^h_{ij}$ then the measure $\mathbb{Q}_i$ of [7] with $f = h$ is recovered.

It is easy to check that the stationary distribution for K^h is given by (3.2.14):

$$\phi(i) = \pi(i)h(i), \quad i \in S.$$

Therefore, the reversal of K^h (with respect to ϕ) is given by

$$\tilde{K}^h_{ij} = \frac{\phi(j)}{\phi(i)}K^h_{ji} = \frac{\pi(j)h(j)}{\pi(i)h(i)}\frac{h(i)}{\alpha h(j)}K_{ji} = \frac{\pi(j)}{\alpha\pi(i)}K_{ji}.$$

That is, $\tilde{K}^h = \frac{1}{\alpha}\tilde{K}$ (it should be clear that $\tilde{K}^h$ is the reversal of K^h with respect to ϕ whereas $\tilde{K}$ is the reversal of K with respect to π described in Section 3.2). This justifies the guess one might have been led to by part a) of Lemma 3.2.3 about how to solve the problem described in the beginning of this section.

Consider the *multiplicative reversiblization* $K^h\left(\frac{1}{\alpha}\tilde{K}\right)$ of K^h with respect to its stationary measure ϕ, and let $\Phi := \mathrm{diag}(\phi)$. If $C := \Phi^{1/2}\left(\frac{1}{\alpha}\tilde{K}\right)\Phi^{-1/2}$ then $K^h\left(\frac{1}{\alpha}\tilde{K}\right) = \Phi^{-1/2}C^\top C\Phi^{1/2}$ is similar to the positive semidefinite matrix $C^\top C$, and so all eigenvalues of $K^h\left(\frac{1}{\alpha}\tilde{K}\right)$ are real and nonnegative. Also, for $\sqrt{\phi} := \mathbf{1}^\top\Phi^{1/2}$, both

$$\sqrt{\phi}C = \phi\left(\frac{1}{\alpha}\right)\tilde{K}\Phi^{-1/2} = \sqrt{\phi}$$

and

$$C\sqrt{\phi}^\top = \Phi^{1/2}\left(\frac{1}{\alpha}\tilde{K}\right)\mathbf{1} = \sqrt{\phi}^\top$$

hold; i.e. $\sqrt{\phi}$ is both a left and right eigenvector for C corresponding to the maximal eigenvalue 1. Let (λ_1, g_1) be the eigenpair for $\tilde{K}$ of Theorem 3.3.2, and assume g_1 can be normalized so that $\|g_1\|_\phi = 1$. Then $v = \Phi^{1/2}g_1$ is an eigenvector for C with eigenvalue λ_1/α and satisfies $\|v\|_2 = 1$. One finds $\sqrt{\phi}v = 0$ by noting $\sqrt{\phi}v = \sqrt{\phi}Cv = (\lambda_1/\alpha)\sqrt{\phi}v$. Let $V_0 = span\{\sqrt{\phi}^\top, v\}$. Suppose that a non-zero spectral gap exists for $K^h\left(\frac{1}{\alpha}\tilde{K}\right)$, and hence that it has a second largest eigenvalue β (this is then the second largest singular value of C). Then by the Courant-Fisher minimax principal for compact operators,

$$\beta = \max_{dim(V)=2}\min_{x\in V,\|x\|_2=1}\|Cx\|_2 \geq \min_{x\in V_0,\|x\|_2=1}\|Cx\|_2 = \|Cv\|_2 = |\lambda_1|/\alpha.$$

Thus

Lemma 4.1.1 *If $K^h\left(\frac{1}{\alpha}\tilde{K}\right)$ has a non-zero spectral gap and if the eigenvector g_1 of Theorem 3.3.2 satisfies $\|g_1\|_\phi < \infty$ then the asymptotic decay rate of χ_n to 0 is $\binom{n}{n(1)-1}\beta^n$ where β is the second largest eigenvalue of $K^h\left(\frac{1}{\alpha}\tilde{K}\right)$.*

4.2 An Accounting Network With Absorption

In this section it is shown with an explicit example that applying the Folk Theorem to substochastic processes on infinite state spaces can present difficulties. This essentially occurs because the kernel $\frac{1}{\alpha}\tilde{K}$ has a tendency to avoid the set of points from which absorption is possible, and this tendency can manifest as a drift away to infinity, precluding arrival at the point z_0 which the large deviation path began from. This transient behavior is stamped out in finite state spaces. The example of this section also provides a demonstration of what can go wrong in trying to apply Theorem 3.3.2.

Consider the following modification to the network of Section 2.4.3: A small insurance firm with an accounting department consisting of a senior accountant and her assistant the junior accountant. Operations in the department proceed as follows: all new claims arrive on the desk of the junior accountant, and do so according to a Poisson process with rate λ. The junior accountant processes the claims at rate μ_2, which subsequently proceed to the senior accountant who carries out the more complicated calculations. The senior accountant processes claims at rate μ_1 at which point they leave the system.

Identify the senior accountant's desk as node 1 and the junior accountant's desk as node 2 in this 2-node network. Let a point (x, y) in the state space $S := \mathbb{N}^2 \setminus \{(x, 0); x \in \mathbb{N}\}$ denote the the number x of claims on the senior accountant's desk and the number y of claims on the junior accountant's desk (see Figure 4.1).

Here the graveyard state consists of all points of the form $(x, 0)$, which means that the department ceases operations if the junior accountant is ever idle (the junior accountant is seen as expendable if such a situation arises and is fired, thereby abolishing the network). Let X_t be the state of the network at time t and let $X = (X_t)_{t\geq 0}$ be the corresponding sub-stochastic Markov chain. Let K be the transition kernel for the embedded discrete time chain $Y = (Y_n)_{n\in\mathbb{N}}$. The transition diagram for Y is

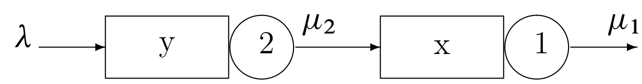

Figure 4.1: Substochastic Accountant Network

given in Figure 4.2.

For the stochastic chain corresponding to an open Jackson network (where the set $\{(x,0); x \in \mathbb{N}\}$ *is* included in the state space) the stationary distribution is given by the well known product formula

$$(x, y) \mapsto (1-\rho_1)\rho_1^x(1-\rho_2)\rho_2^y \tag{4.2.1}$$

where $\rho_1 := \frac{\lambda r_{21}}{\mu_1}$ and $\rho_2 := \frac{\lambda}{\mu_2}$. The sub-stochastic case is not so simple, and the Yaglom limit π, which will be shown to exist, has no product form.

Let α be the eigenvalue corresponding to π, extend π to $S \cup \{(x,0); x \in \mathbb{N}\}$ with the definition $\pi(x,0) := 0$, and let $\pi_x(y) := \sum_{x\geq 0}\pi(x,y)$ and $\pi_y(x) := \sum_{y\geq 1}\pi(x,y)$ denote the marginal densities for π. The equation $\alpha\pi = \pi K$ gives the following interior and boundary conditions:

$$\alpha\pi(x,y) = \lambda\pi(x,y-1) + \mu_1\pi(x+1,y) + \mu_2\pi(x-1,y+1) \tag{4.2.2}$$

$$\alpha\pi(0,y) = \mu_1\pi(0,y) + \lambda\pi(0,y-1) + \mu_1\pi(1,y), \tag{4.2.3}$$

for $x \geq 1, y \geq 1$ (see Figure 4.2).

Summing (4.2.2) over all $x \geq 1$ and employing (4.2.3) gives

$$\alpha(\pi_y(y) - \pi(0,y)) = \lambda(\pi_y(y-1) - \pi(0,y-1)) + \mu_1(\pi_y(y) - \pi(0,y) - \pi(1,y)) + \mu_2\pi_y(y+1)$$

or

$$\alpha\pi_y(y) = \lambda\pi_y(y-1) + \mu_1\pi_y(y) + \mu_2\pi_y(y+1). \tag{4.2.4}$$

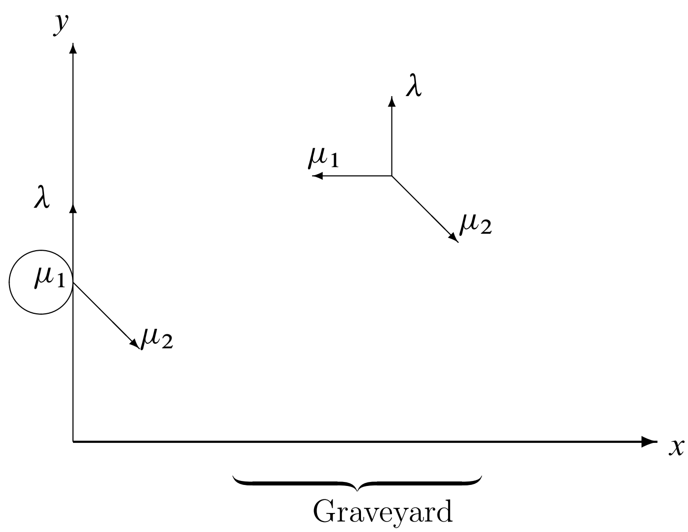

Figure 4.2: Transition probabilities for the accountants' chain

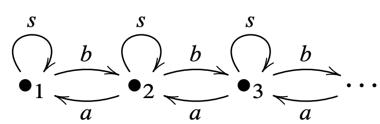

Figure 4.3: Transition Diagram for L^s

If $a := \mu_2$, $b := \lambda$, and $s := \mu_1$, then this means that the marginal π_y corresponds to the one-dimensional sub-stochastic random walk with loops depicted in Figure 4.3.

The transition matrix L^s corresponding to this chain is given by

$$L^s = \begin{bmatrix} s & b & 0 & 0 & \\ a & s & b & 0 & \cdots \\ 0 & a & s & b & \\ & & \ddots & \ddots & \ddots \end{bmatrix} = sI + \underbrace{\begin{bmatrix} 0 & b & 0 & 0 & \\ a & 0 & b & 0 & \cdots \\ 0 & a & 0 & b & \\ & & \ddots & \ddots & \ddots \end{bmatrix}}_{L}. \tag{4.2.5}$$

So if (λ_v, v) is an eigenvalue-eigenvector pair for L then $(s + \lambda_v, v)$ is an eigenvalue-eigenvector pair for L^s. Example 3.1.7 provides the maximal eigenvalue $2\sqrt{ab}$ for L

with its corresponding eigenvector, and therefore

$$\alpha = s + 2\sqrt{ab} \quad \text{and} \quad \pi_y(y) = (1-\sqrt{b/a})^2 y \sqrt{\frac{b}{a}}^{y-1}. \tag{4.2.6}$$

If $\sqrt{ab} < s$ then one can then check that π is given by

$$\pi(x,y) = c\theta_1^x \theta_2^y \left(y + (1-\theta_1^y)\left(x - \frac{\theta_1}{1-\theta_1} \right) \right) \tag{4.2.7}$$

where $\theta_1 := \frac{\sqrt{ab}}{s}$, $\theta_2 := \sqrt{\frac{b}{a}}$, and c is some normalization constant.

Unfortunately, K is not reversible and in fact $\tilde{K}$ has state dependent transitions:

$$\begin{aligned}
\tilde{K}_{(x,y),j} &= (\Pi^{-1} K^\top \Pi)_{(x,y),j} \\
&= \begin{cases} b\theta_2^{-1} \frac{y-1+(1-\theta_1^{y-1})(x-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(x-\theta_1/(1-\theta_1))} & j = (x, y-1) \\ s\theta_1 \frac{y+(1-\theta_1^y)(x+1-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(x-\theta_1/(1-\theta_1))} & j = (x+1, y) \\ a\frac{\theta_2}{\theta_1} \frac{y+1+(1-\theta_1^{y+1})(x-1-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(x-\theta_1/(1-\theta_1))} & j = (x-1, y+1) \end{cases} \\
&= \begin{cases} \sqrt{ab} \frac{y-1+(1-\theta_1^{y-1})(x-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(x-\theta_1/(1-\theta_1))} & j = (x, y-1) \\ \sqrt{ab} \frac{y+(1-\theta_1^y)(x+1-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(x-\theta_1/(1-\theta_1))} & j = (x+1, y) \\ s \frac{y+1+(1-\theta_1^{y+1})(x-1-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(x-\theta_1/(1-\theta_1))} & j = (x-1, y+1) \end{cases}
\end{aligned}$$

for $x > 0, y > 0$ and

$$\tilde{K}_{(0,y),j} = \begin{cases} \sqrt{ab} \frac{y-1+(1-\theta_1^{y-1})(-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(-\theta_1/(1-\theta_1))} & j = (0, y-1) \\ \sqrt{ab} \frac{y+(1-\theta_1^y)(1-\theta_1/(1-\theta_1)))}{y+(1-\theta_1^y)(-\theta_1/(1-\theta_1))} & j = (1, y) \\ s & j = (0, y) \end{cases}$$

for $y > 0$. However, for (x, y) far from the origin the transition rates of $\frac{1}{\alpha}\tilde{K}$ begin to

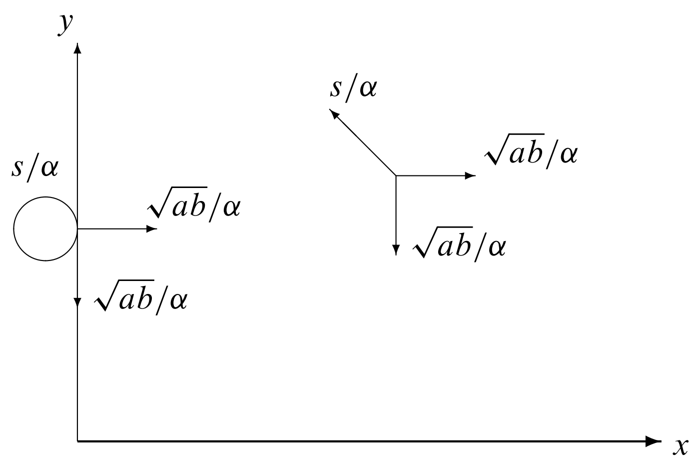

Figure 4.4: Transition Diagram for $\tilde{K}^{\infty}$

approach those of the kernel $\tilde{K}^{\infty}$ on $S \setminus \{(x,0) \mid x \geq 0\}$ given by

$$\tilde{K}^{\infty}_{(x,y),j} := \frac{1}{\alpha}\begin{cases} \sqrt{ab}\,\chi\{y > 1\} & j = (x, y-1) \\ \sqrt{ab}\,\chi\{y = 1\} & j = (x, 1) \\ \sqrt{ab} & j = (x+1, y) \\ s\chi\{x > 0\} & j = (x-1, y+1) \\ s\chi\{x = 0\} & j = (0, y) \end{cases} \tag{4.2.8}$$

(see Figure 4.4).

One rare event of interest for this network is the situation where the desk of the senior account is overflowing with claims. This catastrophe arises as an excursion from $(0,1)$ to a distant point (ℓ, y) (here the point $(\ell, 1)$ is considered). Since ℓ is large such paths require many intermediate steps in which absorption is impossible, and so in light of (4.1.2) K^h most appropriately describes the transition probabilities of such paths as $\ell \to \infty$. Using the methodology described in Section 4.1, one would attempt to observe a chain with initial distribution $I_{(\ell,1)}$ and kernel $\frac{1}{\alpha}\tilde{K}$ until it reaches $(0,1)$. This approach breaks down, however, yielding an interesting conclusion.

First notice that $\tilde{K}^{\infty}$ is a 2-node Jackson network whose flows into and out of node 2 (the y coordinate) are equal. As in Section 2.4.3 artificially enlarge the state space to $S^y = \mathbb{N} \times \mathbb{Z}$ and let $\tilde{K}^{\infty,y}$ be the appropriately modified kernel. The interior fluid limit $\tilde{p}_\infty$ associated with $\tilde{K}^{\infty,y}$ satisfies (2.4.18):

$$\frac{d}{dt}\tilde{p}_\infty(t) = \left(\frac{\sqrt{ab}-s}{\alpha}, \frac{s-\sqrt{ab}}{\alpha}\right),$$

the solution of which is $\tilde{p}_\infty(t) = \tilde{p}_\infty(0) + \frac{(s-\sqrt{ab})}{\alpha} t(-1,1)$. Thus with the initial condition $\tilde{p}_\infty(0) = (1, \frac{1}{\ell})$, $\tilde{p}_\infty$ is a line connecting the points

$$\tilde{p}_\infty(0) = (1,1) \quad \text{and} \quad \tilde{p}_\infty\left(\frac{\alpha}{s-\sqrt{ab}}\right) = (0, 1+1/\ell).$$

Let $T = \frac{\alpha}{s-\sqrt{ab}}$. Thus if the process $\tilde{X}^{\infty}$ is associated with $\tilde{K}^{\infty}$, for any $\epsilon > 0$ and large ℓ one would expect $\tilde{X}$ to lie within the "tube"

$$\tilde{F}_\ell := \left\{ \sup_{0\le t\le T} \left\| \frac{\tilde{X}^{\infty}(\ell t)}{\ell} - \tilde{p}_\infty(t) \right\|_2 \le \epsilon \right\}.$$

This means that paths beginning from $(\ell, 1)$ will make their way toward the axis, hitting it in a region around the point $(0, \ell+1)$.

However, once the backwards process reaches the y axis it is difficult to characterize its behavior. Since the steady state ν for $\tilde{K}^{\infty,y}$ in the x direction is

$$\nu(x) := (1-\theta_1)\,\theta_1^x \quad x \ge 0,$$

the boundary fluid limit $\tilde{p}^1_\infty$ associated with $\tilde{K}^{\infty,y}$ satisfies

$$\frac{d}{dt}\tilde{p}^1_\infty(t) = (1-\theta_1)\,\frac{\sqrt{ab}}{\alpha}(1,-1) + \theta_1\frac{s-\sqrt{ab}}{\alpha}(-1,1) = (0,0)$$

(see (2.4.21)). Thus, the process $\tilde{X}^{\infty}$ has no net drift once it reaches the axis. The situation is worse for $\tilde{X}$ since it can be shown that there is a slight drift in the positive direction, the magnitude of which tends to 0 as $y \to \infty$, but is non negligible at small

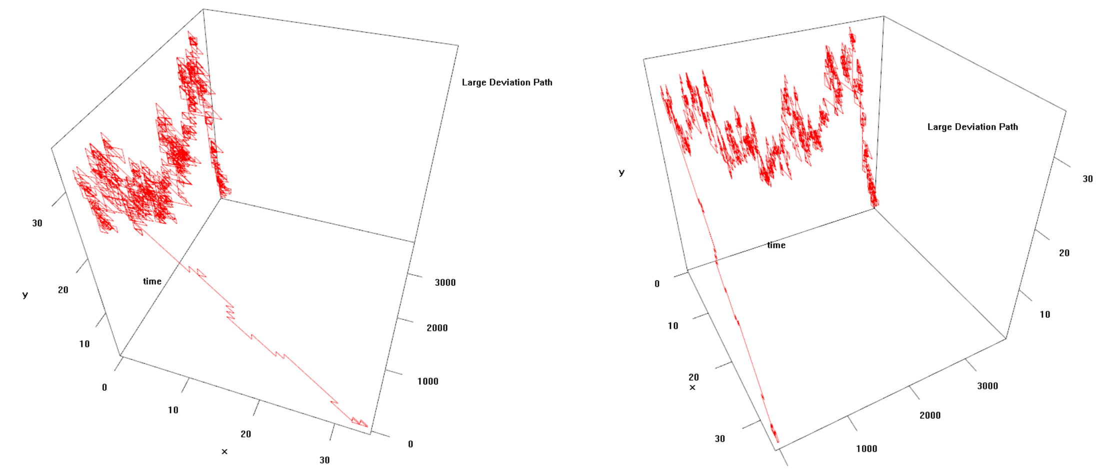

Figure 4.5:]

Backwards Path for Large Deviation

values of y. Therefore there should be no expectation that $\tilde{X}$ will ever reach the point $(0,1)$, and indeed this is verified by simulation.

The process $\tilde{X}^{\infty}$ does eventually arrive at the point $(0,1)$ however, and $\tilde{X}$ can be made to do so by placing reflecting boundaries at $x = B$ and $y = B$ for some value B. Figure 4.5 shows such a path beginning at $(\ell, 1)$ and transitioning to $(0,1)$ for $\ell = 34$, $\lambda = b = 0.1$, $\mu_2 = a = 0.25$, and $s = 0.65$. It can be seen that the process adhered rather closely to the interior fluid limit, and then jittered wildly near the y axis for a time before finally reaching $(0,1)$.

As mentioned at the beginning of this section, it is the transient behavior of $\frac{1}{\alpha}\tilde{K}$ (specifically in the y direction) which prevents some version of the Folk Theorem from holding. Now it will be shown that the hypotheses of Theorem 3.3.2 fail to hold for the marginal process in the y direction. Let $Y^s = \{Y_n^s\}_{n\in\mathbb{N}}$ be a substochastic Markov chain with kernel L^s, and let $Y = \{Y_n\}_{n\in\mathbb{N}}$ be a Markov chain with kernel $\frac{1}{a+b}L$ with L^s, L given by (4.2.5). Figure 4.3 depicts the transitions for Y^s.

Let N_1^s and N_1 be the number of steps, respectively, Y^s and Y make before arriving at state 1. Let U_i^s be iid random variables having geometric density with probability of success $1-s$. The chain Y^s will make a geometrically distributed number of loops (with probability $1-s$ of success) at each state before transitioning to a new state. Among the remaining transitions the proportion which are of the form $i \to i+1$ is a' and the proportion which are of the form $i \to i-1$ is b', and so one can write $N_1^s \stackrel{d}{=} \sum_{i=1}^{N_1}(U_i^s+1)$. Letting $F_{ij}(z)$ and $F_{ij}^s(z)$ be the series (5.1.1) corresponding to $\frac{1}{a+b}L$ and L^s respectively,

$$\begin{aligned} F_{21}^s(z) &= \mathbb{E}_2[z^{N_1^s}] \\ &= \mathbb{E}_2\left[z^{\sum_{i=1}^{N_1}(U_i^s+1)}\right] \\ &= \sum_{k=1}^{\infty} \mathbb{E}_2\left[z^{\sum_{i=1}^{k}(U_i^s+1)}\right] P_2[N_1=k] \\ &= \sum_{k=1}^{\infty} \mathbb{E}_2[z^{U_1^s}]^k z^k P_2[N_1=k]. \end{aligned}$$

Since

$$\mathbb{E}_j[z^{U_1^s}] = \sum_{m=0}^{\infty} z^m s^m (1-s) = \frac{1-s}{1-zs}$$

for any $j \in S$ if $|zs|<1$ it follows that

$$F_{21}^s(z) = \sum_{k=1}^{\infty}\left(\frac{1-s}{1-zs}\right)^k z^k P_2[N_1=k] = F_{21}\left(\frac{(1-s)z}{1-zs}\right). \qquad (4.2.9)$$

Using the formula on pp. 428 of [24] then gives

$$\begin{aligned} F_{21}^s(z) &= \frac{1-\sqrt{1-4a'b'(1-s)^2z^2/(1-zs)^2}}{2b'(1-s)z/(1-zs)} \\ &= \frac{1-\sqrt{1-4abz^2/(1-zs)^2}}{2bz/(1-zs)}. \end{aligned}$$

Since $\sum_{n=1}^{\infty} P_1[N_1^s=n]z^n = zs + \sum_{n=2}^{\infty} P_1[N_1^s=n]z^n = zs + zb\sum_{n=2}^{\infty} P_2[N_1^s = n-1]z^{n-1}$ i.e. $F_{11}^s(z) = zs + zbF_{21}^s(z)$, one can write

$$F_{11}^s(z) = zs + \frac{1-\sqrt{1-4abz^2/(1-zs)^2}}{2/(1-zs)}.$$

Note that

$$\begin{aligned} 1 - 4abz^2/(1-zs)^2 &\geq 0 \quad \Leftrightarrow \\ (1-sz)^2 &\geq 4abz^2 \quad \Leftrightarrow \\ (s^2 - 4ab)z^2 - 2sz + 1 &\geq 0, \end{aligned}$$

which is equivalent to the condition $z \leq \frac{1}{s+2\sqrt{ab}}$ or $z \geq \frac{1}{s-\sqrt{ab}}$. Since $F_{11}(z)$ is an increasing function, it reaches its singularity precisely at

$$R_s := \frac{1}{s + 2\sqrt{ab}}$$

which must therefore be the radius of convergence corresponding to L^s. Note that R_s is, as one might expect, the reciprocal of the eigenvalue given by (4.2.6), and that

$$\begin{aligned} F_{11}(R_s) &= \frac{s}{s+2\sqrt{ab}} + \frac{1}{2/\left(1 - \frac{s}{s+2\sqrt{ab}}\right)} \\ &= \frac{2s}{2(s+2\sqrt{ab})} + \frac{s + 2\sqrt{ab} - s}{2(s+2\sqrt{ab})} \\ &= \frac{s+\sqrt{ab}}{s+2\sqrt{ab}}. \end{aligned}$$

If $s = 0$ this says that $F_{11}(R_s) = \frac{1}{2}$, which is in agreement with [24]. Setting $F_{11}(R_s) = 1$ yields the requirement $ab = 0$, and so there are no values of s for which the matrix L^s is R_s-recurrent. Therefore, Theorem 3.3.2 cannot be applied to find the rate of convergence of χ_n to 0 for the substochastic kernel L^s.

Chapter 5

Finding Quasi-Stationary Measures

In general, finding a quasi-stationary distribution for a substochastic Markov chain can be more difficult than finding the steady state of a fully stochastic chain. The accounting network of Section 4.2 serves as a poignant example of this: the steady state of the corresponding Jackson network is given by the classic formula (4.2.1) whereas the quasi-stationary distribution for the chain is complicated, not easy to arrive at (many thanks to Nicolas Gast for help with this), and only has the nice closed form (4.2.7) by a stroke of luck!

For finite state spaces with small enough cardinality the problem of finding a quasi-stationary distribution is a trivial eigenvector problem. The following sections give techniques for representing or approximating the quasi-stationary distribution when the state space is of large cardinality or infinite. Theorem 5.1.6 can be used to construct the desired measure from a 1-invariant measure of a matrix indexed by a subset A of the state space (in general A may be small, thereby reducing the problem greatly). The kernel of the watched process of Section 2.2 is used as a tool in this construction. Much of what appears in Section 5.1 is motivated by [25], and some of the same ground is covered but with different proofs.

Next, the Approximation Lemma (Lemma 5.2.1) says that the quasi-stationary measure can be constructed as a function of an approximate R^{-1}-invariant measure (with R given in (3.2.17)). The Lemma is an extension of Lemma 5.1 of [8] to the case of substochastic chains. A situation where this result may be useful is one in which the set A from which absorption is possible can be thought of as a kind of boundary, so that restricted to A^c the chain is nice enough that an R^{-1}-invariant measure can be computed. This is the case in Example 5.2.2 where the Approximation Lemma is applied.

5.1 A Representation Theorem

Let $\{Y_n\}_{n\in\mathbb{N}}$ be a discrete time Markov chain with substochastic kernel K and let

$$f_{ij}^{(0)} := 0, \quad f_{ij}^{(n)} := P_i[\tau_j(1) = n], \; n \geq 1,$$

be the probability that the first entry to j occurs on the n^{th} step when starting in state i (here τ_j is given by (2.2.4) with $B = \{j\}$). Let

$$F_{ij}(z) := \sum_{n=1}^{\infty} f_{ij}^{(n)} z^n, \tag{5.1.1}$$

and let R be the radius of convergence of the potential series (3.2.16). Then for any $i, j \in S$

$$\begin{aligned} G_{ij}(z) - \chi\{i = j\} &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \chi\{m \leq n\} f_{ij}^{(m)} z^m K_{jj}^{n-m} z^{n-m} \\ &= \sum_{m=1}^{\infty} f_{ij}^{(m)} z^m \sum_{k=0}^{\infty} K_{jj}^k z^k \\ &= F_{ij}(z) G_{jj}(z); \end{aligned}$$

this is true for any $z \in \mathbb{C}$. For $|z| < R$ these series are convergent (clearly the radius of convergence of $F(z)$ is no less than R) and in particular

$$G_{ii}(z) = \frac{1}{1 - F_{ii}(z)}.$$

If K is R-recurrent then taking $z \uparrow R$ yields ∞ on the left hand side, and hence $F_{ii}(R) = 1$ (see [25] for a more detailed discussion).

For $A \subseteq S$ let $W_n = Y_{\tau_A(n)}$ be the process watched on A, which arises by restricting attention to Y at the entrance times (2.2.4) into A. Lemma 2.2.1 says that W is a Markov chain. Let ${}_AK^{(n)}$ be given by (2.2.5) with $B = A$, let ${}_Af_{ij}^{(0)} = 0$, ${}_Af_{ij}^{(1)} = K_{ij}$, and for $n \geq 2$ define

$$ {}_Af_{ij}^{(n)} := \sum_{k \notin A \cup \{j\}} {}_Af_{ik}^{(n-1)} K_{kj} = P_i[\tau_A(1) \geq n, \tau_j(1) = n]; $$

this is the probability that the chain Y, starting from i, moves to j for the first time via a path that avoids A in the $n-1$ intermediate steps. Note that if $j \in A$ then ${}_Af_{ij}^{(n)} = f_{ij}^{(n)} = {}_AK_{ij}^{(n)}$. Let

$$ {}_AF_{ij}(z) := \sum_{n=1}^{\infty} {}_Af_{ij}^{(n)} z^n, \quad {}_AG_{ij}(z) := \sum_{n=1}^{\infty} {}_AK_{ij}^{(n)} z^n. \tag{5.1.2} $$

Of course, the radii of convergence of these series is no less than R.

Lemma 5.1.1 *For $i \in A$ and $j \in S$,*

$$ {}_AG_{ij}(z) = zK_{ij} + z \sum_{k \notin A} {}_AG_{ik}(z) K_{kj}. $$

Proof: Straightforward calculation yields

$$ \begin{aligned} {}_AG_{ij}(z) &= \sum_{n=1}^{\infty} \sum_{k \notin A} {}_AK_{ik}^{(n-1)} K_{kj} z^n \\ &= zK_{ij} + z \sum_{n=2}^{\infty} \sum_{k \notin A} {}_AK_{ik}^{(n-1)} K_{kj} z^{n-1} \\ &= zK_{ij} + z \sum_{k \notin A} \sum_{n=1}^{\infty} {}_AK_{ik}^{(n)} z^n K_{kj} \\ &= zK_{ij} + z \sum_{k \notin A} {}_AG_{ik}(z) K_{kj}. \end{aligned} $$

∎

For $i, j \in A$ define

$$J_{ij} := {}_A F_{ij}(R) = {}_A G_{ij}(R), \tag{5.1.3}$$

let $\phi_{ij}^0 = \chi\{i = j\}$, $\phi_{ij}^1 := J_{ij}$ and

$$\phi_{ij}^n := \sum_{k \neq j} \phi_{ik}^{n-1} J_{kj}, \quad n \geq 2.$$

This is analogous to the definition of ${}_A f_{ij}^n$, and so ϕ_{ij}^n can be informally thought of as the "probability", starting in i, that the first entry to j occurs on the n^{th} step with kernel J. Strictly speaking however J can be, for instance, superstochastic. Let

$$\Phi_{ij}(z) := \sum_{n=0}^{\infty} \phi_{ij}^n z^n, \quad \Xi_{ij}(z) := \sum_{n=0}^{\infty} J_{ij}^n z^n.$$

Just as before each of the series $\Xi_{ij}(z)$ share a common radius of convergence, and for $i, j \in A$,

$$\begin{aligned} \Xi_{ij}(z) - \chi\{i = j\} &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \chi\{m \leq n\} \phi_{ij}^m z^m J_{jj}^{n-m} z^{n-m} \\ &= \sum_{m=1}^{\infty} \phi_{ij}^m z^m \sum_{k=0}^{\infty} J_{jj}^k z^k \\ &= \Phi_{ij}(z) \Xi_{jj}(z). \end{aligned}$$

In particular,

$$\Xi_{ii}(z) = \frac{1}{1 - \Phi_{ii}(z)} \tag{5.1.4}$$

whenever $|z|$ is less than the radius of convergence of Ξ.

Theorem 5.1.2 *Suppose that a measure π_A on A is invariant for J, i.e.*

$$\pi_A(j) = \sum_{i \in A} \pi_A(i) J_{ij} \text{ for } j \in A.$$

Define the measure π on S via

$$\pi(k) := \sum_{i \in A} \pi_A(i)\, {}_AG_{ik}(R) \text{ for } k \in S.$$

Then $\pi(j) = \pi_A(\ell)$ for all $\ell \in A$ and π is R^{-1}-invariant for K.

Proof: It follows from the definition of π that $\pi(\ell) = \pi_A(\ell)$ if $\ell \in A$. Then for any ℓ,

$$\begin{aligned}
\sum_{k \in S} \pi(k) K_{k\ell} &= \sum_{k \in A} \pi(k) K_{k\ell} + \sum_{k \in A^c} \pi(k) K_{k\ell} \\
&= \sum_{i \in A} \pi_A(i) K_{i\ell} + \sum_{i \in A} \pi_A(i) \sum_{k \in A^c} {}_AG_{ik}(R) K_{k\ell} \\
&= R^{-1} \sum_{i \in A} \pi_A(i) \left(R K_{i\ell} + R \sum_{k \in A^c} {}_AG_{ik}(R) K_{k\ell} \right) \\
&= R^{-1} \sum_{i \in A} \pi_A(i)\, {}_AG_{i\ell}(R) \qquad \text{by Lemma (5.1.1)} \\
&= R^{-1} \pi(\ell).
\end{aligned}$$

∎

Corollary 5.1.3 *If K is R-recurrent then it has an R^{-1}-invariant measure.*

Proof: Let $A = \{i\}$ for any $i \in S$. Then ${}_AF_{ii}(z) = F_{ii}(z)$ and so $\pi_A(j) := \chi\{i = j\}$ satisfies

$$\pi_A(i) J_{ii} = \pi_A(i)\, {}_AF_{ii}(R) = \pi_A(i) F_{ii}(R) = \pi_A(i),$$

i.e. π_A is invariant for J. Then, by Theorem 5.1.2 π_A can be extended to a measure π which is R^{-1}-invariant for K. ∎

Note that Corollary 5.1.3 is implied by Theorem 4.1 of [25], but has been verified here using different techniques.

Lemma 5.1.4 *The matrix J is irreducible, and if K is R-recurrent then J is 1-recurrent.*

Proof: Take arbitrary points $i, j \in A$. For a sequence $i = k_0, k_1, ..., k_m = j$ specifying a path from i to j, i.e. satisfying $K_{k_0k_1} \cdots K_{k_{m-1}k_m} > 0$, let $0 = n_0 < n_1 < ... < n_\ell = m$ be the subsequence of $0, 1, ..., m$ enumerating the points of A (i.e. $k_d \in A$ if and only if $d = n_p$ for some p). If H is the collection of all paths from i to j of length m and having entrances $k_{n_1}, ..., k_{n_\ell} = j$ into A at times $n_1, ..., n_\ell = m$ then

$$
\begin{aligned}
{}_AK^{n_1}_{k_0k_{n_1}} \cdots {}_AK^{m-n_{\ell-1}}_{k_{n_{\ell-1}}k_m} &= \sum_{(j_0,...,j_m)\in H} K_{j_0j_1} \cdots K_{j_{m-1}j_m} \qquad (5.1.5)\\
&\geq K_{k_0k_1} \cdots K_{k_{m-1}k_m}\\
&> 0,
\end{aligned}
$$

(5.1.6)

which means ${}_AK^{n_{r+1}-n_r}_{k_{n_r}k_{n_{r+1}}} > 0$ and thus $J_{k_{n_r}k_{n_{r+1}}} > 0$ for every $r \geq 0$. Then

$$J^\ell_{ij} \geq J_{k_0k_{n_1}} \cdots J_{k_{n_{\ell-1}}k_{n_\ell}} > 0,$$

so J is irreducible.

Now let $i = j$, sum the term in (5.1.5) over all possible values of $k_{n_1}, ..., k_{n_{\ell-1}} \in A$ and every possible collection of ℓ entry times $n_1, ..., n_{\ell-1}, n_\ell = m$, multiply by R^m, then sum over all m. The resulting term on the left is (the rather complicated)

$$
\begin{aligned}
&\sum_{m=1}^{\infty} \sum_{n_1<\cdots<n_\ell=m} \sum_{(k_{n_1},...,k_{n_{\ell-1}})} {}_AK^{n_1}_{ik_{n_1}} R^{n_1} \cdots {}_AK^{m-n_{\ell-1}}_{k_{n_{\ell-1}}i} R^{m-n_{\ell-1}}\\
&= \sum_{(k_{n_1},...,k_{n_{\ell-1}})} \sum_{n_1=1}^{\infty} {}_AK^{n_1}_{ik_{n_1}} R^{n_1} \sum_{n_2=n_1+1}^{\infty} {}_AK^{n_2-n_1}_{k_{n_1}k_{n_2}} R^{n_2-n_1} \cdots \sum_{n_\ell=n_{\ell-1}+1}^{\infty} {}_AK^{n_\ell-n_{\ell-1}}_{k_{n_{\ell-1}}i} R^{n_\ell-n_{\ell-1}}\\
&= \sum_{(j_1,...,j_{\ell-1})\in A^{\ell-1}} \sum_{m_1=1}^{\infty} {}_AK^{m_1}_{ij_1} R^{m_1} \sum_{m_2=1}^{\infty} {}_AK^{m_2}_{j_1j_2} R^{m_2} \cdots \sum_{m_\ell=1}^{\infty} {}_AK^{m_\ell}_{j_{\ell-1}i} R^{m_\ell}
\end{aligned}
$$

$$= \sum_{(j_1,\ldots,j_{\ell-1})\in A^{\ell-1}} J_{ij_1} J_{j_1 j_2} \cdots J_{j_{\ell-1} i}.$$

Now summing over all ℓ gives

$$\sum_{\ell=1}^{\infty} \sum_{(j_1,\ldots,j_{\ell-1})\in A^{\ell-1}} J_{ij_1} J_{j_1 j_2} \cdots J_{j_{\ell-1} i} = \sum_{\ell=1}^{\infty} J_{ii}^{\ell}.$$

This is equivalent to summing (5.1.5) over all paths from i to i of length m, multiplying by R^m, and summing over m, which produces $\sum_{m=0}^{\infty} K_{ii}^m R^m$ on the right side of that equation. This proves that

$$\sum_{m=0}^{\infty} K_{ii}^m R^m = \sum_{\ell=1}^{\infty} J_{ii}^{\ell}.$$

Because of R-recurrence the sum on the left hand side is infinity. Therefore $\sum_{n=0}^{\infty} J_{ii}^n = \infty$. Since $\Phi_{ii}(z)$ is a continuous function of z, the above argument together with (5.1.4) implies that $\Phi_{ii}(1) \geq 1$.

Now suppose that $|z| < 1$. Then

$$\begin{aligned}
\Phi_{ii}(z) &= \sum_{n=1}^{\infty} \sum_{(i_1,\ldots,i_{n-1})\in(A\setminus\{i\})^{n-1}} J_{ii_1} J_{i_1 i_2} \cdots J_{i_{n-1} i} z^n \\
&\leq \sum_{n=1}^{\infty} \sum_{(i_1,\ldots,i_{n-1})\in(A\setminus\{i\})^{n-1}} {}_{\{i\}}F_{ii_1}(R)_{\{i\}}F_{i_1 i_2}(R) \cdots {}_{\{i\}}F_{i_{m-1} i}(R)|z|^n \\
&< \sum_{n=1}^{\infty} \sum_{(i_1,\ldots,i_{n-1})\in(A\setminus\{i\})^{n-1}} {}_{\{i\}}F_{ii_1}(R)_{\{i\}}F_{i_1 i_2}(R) \cdots {}_{\{i\}}F_{i_{m-1} i}(R) \\
&= F_{ii}(R).
\end{aligned}$$

Since $F_{ii}(R) = 1$ this means that $\Phi_{ii}(z) < 1$ for $z < 1$. Since $\Phi_{ii}(1) \geq 1$ continuity of Φ implies that the radius of convergence of Ξ_{ii} is 1. ∎

Corollary 5.1.5 *If K is R-recurrent there are no 1-subinvariant measures for J which are not 1-invariant, and if a 1-invariant measure (such as π_A in Theorem 5.1.2) exists then it is unique up to constant factors.*

Proof: By Lemma 5.1.4 J is 1-recurrent and so this is a direct consequence of Corollary 2 in Section 4 of [25]. ∎

Theorem 5.1.6 *Suppose K is R-recurrent so that (by Corollary 5.1.3) it has an R^{-1}-invariant measure π. Then, for $j \in S$, $\pi(j) = \sum_{i \in A} \pi(i)_A G_{ij}(R)$. In particular, π restricted to A is a 1-invariant measure for J.*

Proof: For $n = 1$, the equation

$$\pi(j) = \sum_{m=1}^{n} \sum_{i \in A} \pi(i)\ {}_A K_{ij}^m R^m + R^n \sum_{i \in A^c} \pi(i)\ {}_A K_{ij}^n \tag{5.1.7}$$

holds by the definition of an R^{-1}-invariant measure. Suppose (5.1.7) holds for some $n \geq 1$. The last term in (5.1.7) is

$$\begin{aligned} R^n \sum_{i \in A^c} \pi(i)_A K_{ij}^n &= R^n \sum_{i \in A^c} \left(R \sum_{k \in S} \pi(k) K_{ki} \right) {}_A K_{ij}^n \\ &= \sum_{k \in A} \pi(k) \sum_{i \in A^c} K_{ki\,A} K_{ij}^n R^{n+1} + R^{n+1} \sum_{k \in A^c} \pi(k) \sum_{i \in A^c} K_{ki}\ {}_A K_{ij}^n \\ &= \sum_{k \in A} \pi(k)\ {}_A K_{kj}^{n+1} R^{n+1} + R^{n+1} \sum_{k \in A^c} \pi(k)\ {}_A K_{kj}^{n+1}, \end{aligned}$$

and so (5.1.7) holds for $n+1$ also. By induction it holds for all n, and letting $n \to \infty$ gives

$$\pi(j) \geq \sum_{i \in A} \pi(i)\ {}_A G_{ij}(R) \text{ for all } j \in S. \tag{5.1.8}$$

Take $j \in A$ so the above implies π restricted to A is 1-subinvariant for J. By Corollary (5.1.5) any subinvariant measure must in fact be invariant so π restricted to A is 1-invariant for J. Hence by Theorem 5.1.2, defining

$$\sigma(j) = \sum_{i \in A} \pi(i)_A G_{ij}(R)$$

for $j \in S$ yields an R^{-1}-invariant measure for K, and by Corollary 2 in Section 4 of [25] σ is the unique measure with this property, up to constant factors. Since π has this property and agrees with σ on A, $\pi = \sigma$ and the result follows. ∎

Corollary 5.1.7 *Any two R^{-1}-invariant measures π and σ for an R-recurrent kernel K are constant multiples of each other.*

Proof: Let $A = \{i\}$. Without loss of generality assume that each measure is normalized so that $\pi(A) = \sigma(A) = 1$. Hence, by Theorem 5.1.6, $\pi(j) = {}_AG_{ij}(R)$ and $\sigma(j) = {}_AG_{ij}(R)$. In other words, before normalization σ and π are multiples. ∎

This is again a result which is implied by Corollary 2 of Section 4 of [25].

To see the power of Theorem 5.1.6, suppose an R-recurrent kernel K is in hand for which the unique quasi-invariant measure π is sought. Examples of interest are those in which the state space is infinite or at least of large enough cardinality as to make computerized eigenvalue/eigenvector analyses intractable. Perhaps K exhibits some degree of "regularity" across much of the state space but is irregular with respect to transitions into or out of a set A of manageable size. Then the procedure for finding π suggested by Theorem 5.1.6 is to form the small $|A| \times |A|$ matrix J and compute its unique 1-invariant measure π_A, a much simpler computation, and then compute π as

$$\pi(j) = \sum_{i \in A} \pi(i)\, {}_AG_{ij}(R).$$

5.2 The Approximation Lemma

Suppose an irreducible R-recurrent kernel K has R^{-1}-invariant measure π, and suppose one can find an *approximate* R^{-1}-invariant measure ψ for K, i.e. satisfying, for some set $A \subseteq S$,

$$R^{-1}\psi(i) = \sum_{k\in S} \psi(k)K_{ki} \quad \text{for } i \in A^c. \tag{5.2.1}$$

Define the *approximate time reversal* of K with respect to ψ to be the matrix

$$\overleftarrow{K}_{ij} := \frac{R\psi(j)}{\psi(i)} K_{ji}. \tag{5.2.2}$$

Note that if $i \in A^c$ then $\sum_{j\in S} \overleftarrow{K}_{ij} = 1$. Thus it is possible to define a sub-stochastic Markov chain $\overleftarrow{Y} = \{\overleftarrow{Y}_n\}_{n\in\mathbb{N}}$ which transitions from state $i \in A^c$ to $j \in A^c$ with probability $\overleftarrow{K}_{ij}$ and is absorbed on the set A, starting from $i \in A^c$, with probability $\sum_{j\in A} \overleftarrow{K}_{ij}$. Let $\overleftarrow{\tau} = \inf\{n \geq 0 : \overleftarrow{Y}_n \in A\}$, which may be infinite.

Lemma 5.2.1 *For $j \in A^c$,*

$$\frac{\pi(j)}{\psi(j)} = \mathbb{E}_j \left[\frac{\pi(\overleftarrow{Y}_{\overleftarrow{\tau}}))}{\psi(\overleftarrow{Y}_{\overleftarrow{\tau}})} \chi\{\overleftarrow{\tau} < \infty\} \right]. \tag{5.2.3}$$

Proof: Equation (5.2.2) implies that

$$\begin{aligned}
{}_AK^n_{ij} &= P_i[Y_1 \notin A, \ldots, Y_{n-1} \notin A, Y_n = j] \\
&= P_j[\overleftarrow{Y}_1 \notin A, \ldots, \overleftarrow{Y}_{n-1} \notin A, \overleftarrow{Y}_n = i] \\
&= \frac{\psi(j)}{\psi(i)R^n} {}_A\overleftarrow{K}^n_{ji}.
\end{aligned}$$

According to Theorem (5.1.6), $j \in A^c$ implies

$$\begin{aligned}
\pi(j) &= \sum_{i\in A} \pi(i) {}_AG_{ij}(R) \\
&= \sum_{i\in A} \pi(i) \sum_{n=1}^{\infty} {}_AK^n_{ij} R^n \\
&= \sum_{i\in A} \pi(i) \sum_{n=1}^{\infty} \frac{\psi(j)}{\psi(i)} {}_A\overleftarrow{K}^n_{ji} \\
&= \psi(j) \sum_{n=1}^{\infty} \sum_{i\in A} {}_A\overleftarrow{K}^n_{ji} \frac{\pi(i)}{\psi(i)}
\end{aligned}$$

$$= \psi(j)E_j\left[\frac{\pi(\overleftarrow{Y}_{\overleftarrow{\tau}})}{\psi(\overleftarrow{Y}_{\overleftarrow{\tau}})}\chi\{\overleftarrow{\tau} < \infty\}\right],$$

which establishes (5.2.3). ∎

Example 5.2.2 *Let $0 < \lambda < \mu < 0$ be such that $\lambda = 1 - \mu$ and let K be the irreducible, substochastic kernel on $S := \mathbb{N}$ given by*

$$K_{ij} := \begin{cases} \lambda & j = i+1 \\ \mu & j = i-1 \neq 0, \quad or \quad i = j = 0\,; \\ \mu_0 & i = 1, j = 0 \end{cases}$$

here $\mu_0 \in (0, \mu)$ is the probability of surviving a jump from state 1 to state 0. Let X be a process whose transition probabilities are described by K. In [24] it was determined that $F_{21}(z) = \frac{1-\sqrt{1-4\lambda\mu z^2}}{2\lambda z}$, and since $f_{01}^{(k)} = \mu^{k-1}\lambda$ it follows that

$$\begin{aligned} F_{11}(z) &= \sum_{n=1}^{\infty} f_{10}^{(1)} f_{01}^{(n-1)} z^n + \sum_{n=1}^{\infty} f_{12}^{(1)} f_{21}^{(n-1)} z^n \\ &= \mu_0 z F_{01}(z) + \lambda z F_{21}(z) \\ &= \mu_0 z \sum_{n=1}^{\infty} \mu^{n-1}\lambda z^n + \frac{1-\sqrt{1-4\lambda\mu z^2}}{2} \\ &= \lambda\mu_0 z^2 \frac{1}{1-\mu z} + \frac{1-\sqrt{1-4\lambda\mu z^2}}{2} \end{aligned}$$

for $|z| < \min\left\{\frac{1}{\mu}, \frac{1}{2\sqrt{\lambda\mu}}\right\}$. This minimum is equal to $\frac{1}{\mu}$ precisely when

$$2\sqrt{\lambda\mu} < \mu, \qquad or \qquad 4\lambda < \mu. \tag{5.2.4}$$

Assume that (5.2.4) holds. Then on $[0, \infty)$ the function $F_{11}(z)$ is non-decreasing and encounters a singularity at $z = 1/\mu > 1$. Recall $G_{ii}(z) = \frac{1}{1-F_{ii}(z)}$ for $|z| < R$ where

R is the radius of convergence of $G_{ii}(z)$. Thus $R \leq \frac{1}{\mu}$. Setting $F_{11}(R) = 1$ yields

$$\begin{aligned} 2(1-\mu z) &= 2\lambda\mu_0 z^2 + (1-\mu z)(1-\sqrt{1-4\lambda\mu z^2}) && \iff \\ 1-\mu z - 2\lambda\mu_0 z^2 &= -(1-\mu z)(\sqrt{1-4\lambda\mu z^2}) && \implies \\ (1-\mu z-2\lambda\mu_0 z^2)^2 &= (1-\mu z)^2(1-4\lambda\mu z^2). \end{aligned}$$

After some simplification this reduces to the polynomial equation

$$0 = (\mu_0 - \mu) + \mu(2\mu - \mu_0)z - (\mu^3 + \lambda\mu_0^2)z^2.$$

This will have a real solution $R \in \left(1, \frac{1}{\mu}\right)$ precisely when K is R-recurrent. The subset of the parameter space where this occurs is non-empty; for example taking $\mu = 0.6$ and $\mu_0 = 0.5$ yields $R \approx 1.0184$. Let $R > 1$ be the root of this polynomial which gives the radius of convergence of $G_{ii}(z)$. Since $F_{11}(R) = 1$ the kernel K is R-recurrent. Let $A = \{0, 1\}$ and

$$J = \begin{bmatrix} F_{00}(R) & F_{01}(R) \\ F_{10}(R) & {}_AF_{11}(R) \end{bmatrix} = \begin{bmatrix} \mu R & \lambda R \\ \mu_0 R & \frac{1-\sqrt{1-4\lambda\mu R^2}}{2} \end{bmatrix}.$$

By Lemma 5.1.4 this matrix has 1 as an eigenvalue. Let $B = \{0, 1, 2, 3\}$ and consider that, since the transition rates into and out of each state $i > 1$ are equal, ${}_BF_{33}(z) = {}_AF_{11}(z)$. Therefore

$$\begin{aligned} F_{31}(z) &= \mu^2 z^2 + \lambda\mu z^2 F_{31}(z) + {}_BF_{33}(z)F_{31}(z) \\ &= \mu^2 z^2 + \lambda\mu z^2 F_{31}(z) + {}_AF_{11}(z)F_{31}(z), \end{aligned}$$

which means that

$$F_{31}(z) = \frac{\mu^2 z^2}{1 - \lambda\mu z^2 - {}_AF_{11}(z)}.$$

Now

$$\begin{aligned} {}_AF_{11}(z) &= \lambda\mu z^2 + \sum_{n=3}^{\infty} f_{11}^{(n)} z^n \\ &= \lambda\mu z^2 + \lambda^2 z^2 F_{31}(z) \end{aligned}$$

$$= \lambda\mu z^2 + \lambda^2 z^2 \frac{\mu^2 z^2}{1 - \lambda\mu z^2 - {}_A F_{11}(z)},$$

and multiplying both sides of this equation by $1 - \lambda\mu z^2 - {}_A F_{11}(z)$ *yields*

$$(1 - \lambda\mu z^2 - {}_A F_{11}(z))\, {}_A F_{11}(z) = \lambda\mu z^2 - \lambda\mu z^2\, {}_A F_{11}(z).$$

Therefore

$${}_A F_{11}(z) = \lambda\mu z^2 + {}_A F_{11}(z)^2. \tag{5.2.5}$$

Setting $z = R$*, dividing both sides of this equation by* $R^2\mu$ *and letting* $\rho := \frac{1}{R\mu}\, {}_A F_{11}(R)$ *gives*

$$\frac{1}{R}\rho = \lambda + \mu\rho^2.$$

This means that $\psi(i) := (1-\rho)\rho^i$ *satisfies (5.2.1) and hence is an approximate* R^{-1}*-invariant measure for* K*. Thus the Approximation Lemma can be applied. Since* $\overleftarrow{Y}_{\overleftarrow{\tau}} \equiv 1$*, the Approximation Lemma says that the quasi-stationary measure* π *for* K *satisfies*

$$\pi(j) = \frac{\pi(1)}{\psi(1)} P_j[\overleftarrow{\tau} < \infty]\psi(j), \quad j \in A^c. \tag{5.2.6}$$

Note that

$$\overleftarrow{K}_{ij} = \frac{R\psi(j)}{\psi(i)} K_{ji} = \begin{cases} \frac{\lambda\mu R^2}{{}_A F_{11}(R)} & j = i-1 \\ {}_A F_{11}(R) & j = i+1 \end{cases}$$

for $i \in A^c$*. Let* $\overleftarrow{\lambda} := {}_A F_{11}(R)$ *and* $\overleftarrow{\mu} := \frac{\lambda\mu R^2}{{}_A F_{11}(R)}$*. Note that by (5.2.5)*

$$\overleftarrow{\lambda} + \overleftarrow{\mu} = \frac{{}_A F_{11}(R)^2 + \lambda\mu R^2}{{}_A F_{11}(R)} = 1,$$

as required. Note also that (5.2.4) and $R < \frac{1}{\mu}$ *imply* $0 < 4\lambda\mu R^2 < 1$*. Then*

$$\begin{aligned} 1 - 4\lambda\mu R^2 &< \sqrt{1 - 4\lambda\mu R^2} && \Longrightarrow \\ (1 - \sqrt{1 - 4\lambda\mu R^2})^2 &< 4\lambda\mu R^2 && \Longrightarrow \\ {}_A F_{11}(R)^2 &< \lambda\mu R^2 && \Longrightarrow \\ \overleftarrow{\lambda} &< \overleftarrow{\mu}. \end{aligned}$$

Therefore $P_j[\overleftarrow{\tau} < \infty] = 1$ *and (5.2.6) becomes*

$$\pi(j) = \pi(1)\rho^{j-1}, \quad j > 1.$$

Chapter 6

Conclusions and Future Work

The aim of this thesis has been to make some constructive observations about different varieties of rare events.

For stationary stochastic processes there is the Folk Theorem (Theorem 2.4.2) which explains nicely what one should expect the typical large deviation path to look like, under suitable conditions. One needs only to have an understanding of how the reversed process evolves, which is always true when dealing with reversible and stationary Markov chains, and a "tube" of backward trajectories $\tilde{\boldsymbol{F}}_\ell$ within which the reversed process remains asymptotically. This theorem successfully plants with rigorous footing the time reversal techniques which have already been in use for some time, in a setting (stationary point processes) which is more general than the Markovian case.

The specialization of this result to discrete time Markov chains in Theorem 2.4.6 links the abstract to the checkable, and the accountants example of Section 2.4.3 demonstrates nicely how the Folk Theorem can be used to find the most likely path to a snowed in senior accountant. This could allow the senior accountant to forestall a catastrophe by keeping an eye out for early warning signs that the large deviation path is underway.

The extension of these reversal techniques to the substochastic Markovian case is not complete. One problem is the lack of stationarity for substochastic Markov processes. If a new Folk Theorem is to be formulated which does not require stationarity then it will likely bear little resemblance to Theorem 2.4.2, and in any case it is not clear how this would be done. Instead, it was argued in Section 4 that one should study a fully stochastic transformation of the substochastic kernel (namely the sustained kernel K^h given by (4.1.2)). There are good reasons to consider this kernel, but in the R-transient case it is difficult to describe the tube of backwards trajectories $\tilde{F}_\ell$ originating from the rare event with this device alone. Some new techniques will need to be developed in a future work.

In the case that a unique Yaglom limit exists for a substochastic chain, it plays a prominent role in the formation of a kernel $\frac{1}{\alpha}\tilde{K}$ governing the backwards process of the chain starting at a rare initial point. Thus in Chapter 3 this important object (the Yaglom limit π) was introduced and a brief review of some standard results was given. It is natural to ask how quickly the Yaglom limit is reached, and under some conditions on the initial distribution π_0 and the assumption of R-positivity, this question was successfully answered in Theorem 3.3.2. It was also demonstrated that the straightforward spectral analysis which gives the rate of convergence in the fully stochastic setting can fail to produce correct conclusions in the substochastic setting.

The conditions on π_0 required to prove Theorem 3.3.2 are known to hold on all finite state spaces, but the infinite dimensional case is not fully understood. It may be that the infinite dimensional case is very sensitive to the metric which measures the distance from the conditional distribution π_n at time n and the Yaglom limit π. This is a difficult topic, but the humble hope is that more headway can be made in the future.

Finally, in Chapter 5 the problem of finding quasi-stationary measures (or Yaglom limits) is addressed in two main results: the Representation Theorem and the Approximation Lemma. The first is an extension of the work [25] of Vere-Jones, and the second is an extension of Lemma 5.1 of [8] to the substochastic case. It was shown in Example 5.2.2 that these techniques can work nicely to yield quasi-stationary distributions.

Bibliography

[1] I. J. B. F. Adan, R. Foley, D. McDonald, (2009). *Exact asymptotics for the stationary distribution of a Markov chain: a production model*. Queueing Syst., 311–344.

[2] V. Anantharam, P. Heidelberger, and P. Tsoucas, (1990). *Analysis of Rare Events in Continuous Time Markov Chains via Time Reversal and Fluid Approximation* . Research Report, Computer Science Division, IBM T.J. Watson Center.

[3] F Baccelli, P. Bremaud, (2003). "Elements of Queueing Theory" Springer Verlag.

[4] F Baccelli, D McDonald,(1997). *Rare events for stationary processes*. Stochastic Processes and their Applications **89**, 141–173.

[5] P. Billingsley, (1995). "Probability and Measure" Wiley Series in Probability and Mathematical Statistics.

[6] A.A. Borovkov, A.A. Mogul'skii, (2001). *Limit theorems in the boundary hitting problem for a multidimensional random walk*. Siberian Mathematical Journal **4**, 245–270.

[7] L.A. Breyer, G.O. Roberts (1997). *A Quasi-Ergodic Theorem for Evanescent Processes*, preprint.

[8] J. Collingwood, R. Foley and D. McDonald, (2012). *Networks with Cascading Overloads*, Journal of Industrial and Management Optimization, **4**, 877–894.

[9] P. Dupuis, R. S. Ellis, A. Weiss, (1991). *Large Deviations for Markov Processes with Discontinuous Statistics, I: General Upper Bounds*. The Annals of Probability, Vol. 19, No. 3, 1280 – 1297.

[10] P. A. Ferrari, H. Kesten, S. Martinez, P. Picco, (1995). *Existence of Quasi-Stationary Distributions. A Renewal Dynamical Approach*. The Annals of Probability, Vol. 23, No. 2, 501 – 521.

[11] J. A. Fill, (1991). *Eigenvalue Bounds on Convergence to Stationarity for Non-reversible Markov Chains, With an Application to the Exclusion Process*. The Annals of Applied Probability, Vol. 1, No. 1, 62–87.

[12] R. Foley, D. McDonald, (2001). *Join the shortest queue: stability and exact asymptotics*, Annals of Applied Probability, **11**, 569–607.

[13] R. Foley, D. McDonald, (2004). *Large deviations of a modified jackson network: stability and rough asymptotics*, Annals of Applied Probability, **15**, 519–541.

[14] R. Foley, D. McDonald, (2005). *Bridges and Networks: Exact Asymptotics*, *Annals of Applied Probability,* **15**, pp. 542–586.

[15] R. Foley, D. McDonald, *Constructing a harmonic function for an irreducible non-negative matrix with convergence parameter* $R > 1$. Accepted subject to revisions London Mathematical Society.

[16] I. Ignatiouk-Robert, C. Loree,(2010). *Martin boundary of a killed random walk on a quadrant* Ann. Probab. **38** , 1106–1142.

[17] H. Kesten, (1995). *A Ratio Limit Theorem For (Sub) Markov Chains on* $\{1, 2, ...\}$ *With Bounded Jumps*. Adv. Appl. Prob. **27**, 652–691.

[18] T. Kurtz, (1978). *Strong approximation theorems for density dependent Markov chains*, Stochastic Processes and their Applications **6**, 223–240.

[19] K. Majewski, K. Ramanan,(2008). *How large queue lengths build up in a Jackson network*, preprint.

[20] S.P. Meyn, R.L. Tweedie, (1993). *Markov chains and stochastic stability. Springer-Verlag*, London. Available at: probability.ca/MT.

[21] M. Miyazawa, Y. Zhao, (2004). *The stationary tail asymptotics in the GI/G/1-type queue with countably many background states* Advances in Applied Probability, **36**, 1231 – 1251.

[22] V. Nicola, T. Zaburnenko, (2007).*Efficient importance sampling heuristics for the simulation of population overflow in jackson networks.* ACM Transactions on Modeling and Computer Simulation, **17**.

[23] A. Schwartz, A. Weiss, (1995). "Large Deviations For Performance Analysis" Chapman & Hall.

[24] E. Seneta, D. Vere-Jones, (1966). *On Quasi-Stationary Distributions in Discrete-time Markov Chains with a Denumerable Infinity of States.* Journal of Applied Probability, Vol. 3, No. 2, 403–434.

[25] D. Vere-Jones (1967). *E*rgodic Properties of Non-Negative Matrices I. *Pacific Journal of Mathematics* **22**, 361-386.