The Fundamental Limit Theorem of Countable Markov Chains

Nathanael Gentry

A Senior Thesis submitted in partial fulfillment
of the requirements for graduation
in the Honors Program
Liberty University
Spring 2021

Acceptance of Senior Honors Thesis

This Senior Honors Thesis is accepted in partial fulfillment of the requirements for graduation from the Honors Program of Liberty University.

____
David E. Schweitzer, Ph.D.
Thesis Chair

____
Daniel S. Joseph, Ph.D.
Committee Member

____
Cynthia A. Goodrich, Ed.D.
Assistant Honors Director

____
Date

Abstract

In 1906, the Russian probabilist A.A. Markov proved that the independence of a sequence of random variables is not a necessary condition for a law of large numbers to exist on that sequence. Markov's sequences – today known as Markov chains – touch several deep results in dynamical systems theory and have found wide application in bibliometrics, linguistics, artificial intelligence, and statistical mechanics. After developing the appropriate background, we prove a modern formulation of the law of large numbers (fundamental theorem) for simple countable Markov chains and develop an elementary notion of ergodicity. Then, we apply these chain convergence results to study PageRank and the Google matrix.

Keywords: Markov chains, stochastic processes, dynamical systems, law of large numbers.

The Fundamental Limit Theorem of Countable Markov Chains

The law of large numbers dates to the very nascence of probability theory, with the posthumous *Ars Conjectandi* of Jacob Bernoulli (1713). Published when Bayes was just 11 years old, Bernoulli's thoughts on convergence were threaded through mathematical discourse by many pioneers of probability (Seneta, 2013). In its modern statement, we define a sequence of independent and identically distributed random variables (X_n). Then endow each element X_n with finite mean $\mathbb{E}(X_n) := \mu$. In the *strong law of large numbers*, the sample mean $\overline{X_n}$ converges *almost surely* to μ. That is, the set of exceptions might not be empty, but it carries no probability mass:

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{X_n} \neq \mu\right) = 0\,. \tag{1}$$

Sequences of independent and identically distributed random variables are a special case of *stationary sequences* (Fristedt & Gray, 1997), and a rich theory has developed around the asymptotic properties of such sequences. This thesis proves a law of large numbers (fundamental limit theorem) for a special stationary sequence – the positive recurrent Markov chain.

During the time of A.A. Markov (1856-1922), probabilists were attempting to generalize the conditions under which a law of large numbers would exist on a sequence. In 1902, the Russian probabalist P.A. Nekrasov (1853-1924) asserted that pairwise independence of the (X_n) was necessary for the law of large numbers; there could be no futher generalization. However, Markov – one of Chebyshev's distinguished disciples – detested Nekrasov and endeavored to prove him wrong (Seneta, 1996). In his famous 1906 paper, Markov incisively constructed a counterexample for Nekrasov.

A modern formulation of this counterexample considers a surfer on a network that has only two pages, p_1 and p_2. Each page has a single link on it. Every so often, the surfer clicks the link on the current page. Suppose the link on page p_1 has probability p of sending the surfer to p_2, and probability $(1 - p)$ of sending the surfer back to p_1. Let q be the analogous probability for the link on p_2. To track the surfer through time, we can construct a random variable $\mathcal{X}_t$ for each timestep t. Each of these random variables, valued over p_1 and p_2, gives the probability that the surfer is on that page during the given time.

Unlike the classic coin-flipping experiment, this demonstration does not quite consist of independent and identically distributed trials. Because the probabilities of transition are not necessarily equal for each page, the probability of landing on p_1 or p_2 in the next timestep depends upon the page currently visited. Markov (1951) thus concluded, "independence of quantities does not constitute a necessary condition for the law of large numbers" (p. 507).

Basic Theory

Markov originally studied *finite* chains – those with a finite state space. A theory of *countable* chains waited until 1936, with a paper of Komolgorov (Derman, 1955). Around this time, more than a decade after his death, Markov's chains were first named in his honor (Seneta, 2006). Markov's creation resembled the modern form of the *Markov chain*, the modern mathematical object that bears his name. A *discrete-time Markov chain* $\mathcal{M} := \big((\mathcal{X}_n), \mathcal{S}, \mathcal{P}\big)$ is a discrete sequence of random variables $(\mathcal{X}_n)$ on a countable state space $\mathcal{S}$ such that for all states $s \in \mathcal{S}$ over all times t, the *Markov property* holds:

$$\mathbb{P}\left(\mathcal{X}_t = s_t \mid \mathcal{X}_i = s_i,\ 0 \leq i < t\right) = \mathbb{P}\left(\mathcal{X}_t = s_t \mid \mathcal{X}_{t-1} = s_{t-1}\right)\ . \tag{2}$$

A *homogeneous* chain is described by a *transition matrix* $\mathcal{P} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$, where for all t,

$$(\mathcal{P})_{i,j} := \mathbb{P}\left(\mathcal{X}_{t+1} = s_j \mid \mathcal{X}_t = s_i\right) . \tag{3}$$

A Markov chain selects a new state at each timestep – even if the current state is rechosen. Thus, for a state s_j over all times t,

$$\sum_{s_i \in \mathcal{S}} \mathbb{P}\left(\mathcal{X}_{t+1} = s_i \mid \mathcal{X}_t = s_j\right) = 1 . \tag{4}$$

Equivalently, the rows $\mathcal{P}$ must sum to unity, thus ensuring that $\mathcal{P}$ is a *(row) stochastic* matrix.

Since the Perron-Frobenius theory of nonnegative matrices developed alongside Markov's theory of chains (Seneta, 2006), his seminal papers did not use concepts of recurrence, irreducibility, and periodicity – the major ideas in this thesis. Although Markov did not rigorously address the convergence concerns thus imposed, he eventually anticipated much of the requisite matrix theory in 1908 – four years before Frobenius's landmark paper (Schneider, 1977). In this thesis, we take a probabilistic approach to the fundamental theorem (law of large numbers) for Markov chains, but we will retrofit the useful graph-theoretical structure developed since Markov.

Transition Matrices

Unless stated otherwise, henceforth let a homogeneous discrete-time Markov chain $\left((\mathcal{X}_n), \mathcal{S}, \mathcal{P}\right)$ be given. For all times t, the k^{th} *transition matrix* $\mathcal{P}^{(k)} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ is given by

$$\left(\mathcal{P}^{(k)}\right)_{i,j} := \mathbb{P}\left(\mathcal{X}_{t+k} = s_j \mid \mathcal{X}_t = s_i\right) . \tag{5}$$

The well-known Chapman-Kolmogorov equations, named after the two pioneering probabilists who independently discovered them, finds its elegant expression through the language of matrix multiplication. For a transition matrix $\mathcal{P}$ and some times t and t',

$$\left(\mathcal{P}^{(t+t')}\right)_{i,j} = \left(\mathcal{P}^{t}\mathcal{P}^{t'}\right)_{i,j} = \left(\mathcal{P}^{t+t'}\right)_{i,j} . \tag{6}$$

Dynkin (1989) examines rigorous proofs of this elementary fact, which implies we have $\mathcal{P}^{(0)}$ as the identity matrix. In other words, for a state s_j, we have the tautology

$$\mathbb{P}\left(X_t = s_j \mid X_t = s_j\right) = 1 . \tag{7}$$

Return to our simple Web surfer, which has transition matrix

$$\mathcal{P} := \begin{bmatrix} \mathcal{P}_{1,1} & \mathcal{P}_{1,2} \\ \mathcal{P}_{2,1} & \mathcal{P}_{2,2} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} . \tag{8}$$

As shown in Figure 1, the transition matrix can be viewed as a directed graph, where the weight of the edge between states s_i and s_j is simply $(\mathcal{P})_{i,j}$. Understanding a chain as a graph produces a useful classification of states. State s_j is *reachable from* state s_i if there is some time $t > 0$ whereby $\left(\mathcal{P}^t\right)_{i,j} > 0$. This is a directed t-path of positive probability from s_i to s_j, so

$$\sum_{t=0}^{\infty} \mathbb{P}\left(X_t = s_j \mid X_0 = s_i\right) > 0 . \tag{9}$$

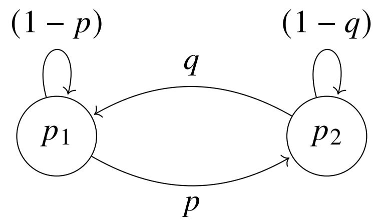
This sum will become more useful later.

Communication of States

States s_i and s_j *communicate* if each is reachable from the other. Communication of states forms an equivalence relation. Since $(\mathcal{P}^0)_{j,j} = 1$ for all states, communication is reflexive. Symmetry follows directly from the definition. For transitivity, let states s_i and s_j communicate, and let s_j and s_k communicate. So there is a time t such that $\left(\mathcal{P}^t\right)_{i,j} > 0$ and a time t' such that $\left(\mathcal{P}^{t'}\right)_{j,k} > 0$. Note that $\left(\mathcal{P}^{t+t'}\right)_{i,k}$ gives the total probability of transition from s_i to s_k across all

Figure 1

Transition Graph for Random Web Surfer



$(t + t')$-paths, and so this probability must be at least the probability of such a transition by a more specific path – specifically, the path that first visits s_j. Thus,

$$\left(\mathcal{P}^{t+t'}\right)_{i,k} \geq \left(\mathcal{P}^{t}\right)_{i,j} \left(\mathcal{P}^{t'}\right)_{j,k} > 0\,. \tag{10}$$

Reversing the order of s_i and s_k finishes the argument for transitivity, and thus proves that these two states comunicate. Because of this equivalence relation, a chain's state space can be partitioned into classes of communicating states.

Reducibility

An important class of infinite-state examples in Markov theory are the *random walks* (Levin, Peres, & Wilmer, 2008), and we consider several forms of the *simple* (one-dimensional) random walk. Such a random walk is a Markov chain with states $s \in \mathbb{Z}$ where

$$\mathbb{P}\left(X_{t+1} = s + 1 \mid X_t = s\right) := p\,, \tag{11}$$

$$\mathbb{P}\left(X_{t+1} = s - 1 \mid X_t = s\right) := 1 - p \tag{12}$$

for some probability p, with zeros elsewhere. The *symmetric* simple random walk has $p = \frac{1}{2}$. Simple random walks can have *reflecting boundaries* – states with transitions like the above but with endpoints $\alpha, \beta \in \mathbb{Z}$:

$$\mathbb{P}\left(\mathcal{X}_{t+1} = \alpha + 1 \mid \mathcal{X}_t = \alpha\right) := p_\alpha\,, \tag{13}$$

$$\mathbb{P}\left(\mathcal{X}_{t+1} = \beta - 1 \mid \mathcal{X}_t = \beta\right) := p_\beta\,. \tag{14}$$

Figure 2 illustrates these situations. When $p_\alpha = p_\beta = 1$, the chain has *absorbing boundaries*. These random walks are *reducible*: They can be broken into disjoint communicating classes – a trivial (absorbing) class for each absorbing state, and a reflecting boundary walk for the inner states. An *irreducible* chain, on the other hand, has a strongly connected graph. The simple random walk is irreducible, since

$$\mathbb{P}\left(\mathcal{X}_{|s_i - s_j|} = s_i \,\middle|\, \mathcal{X}_0 = s_j\right) = p^{|s_i - s_j|}\,. \tag{15}$$

Periodicity

Note that once a state is visited in a simple random walk, that state can only be visited again in a multiple of two timesteps from the current time. In general, let s_j be a state in an arbitrary chain, and define the *period* of s_j as the greatest common divisor of the *possible return times*:

$$\gamma_j := \gcd\left(\left\{n \geq 1 : \mathbb{P}\left(\mathcal{X}_n = s_j \mid \mathcal{X}_0 = s_j\right) > 0\right\}\right), \tag{16}$$

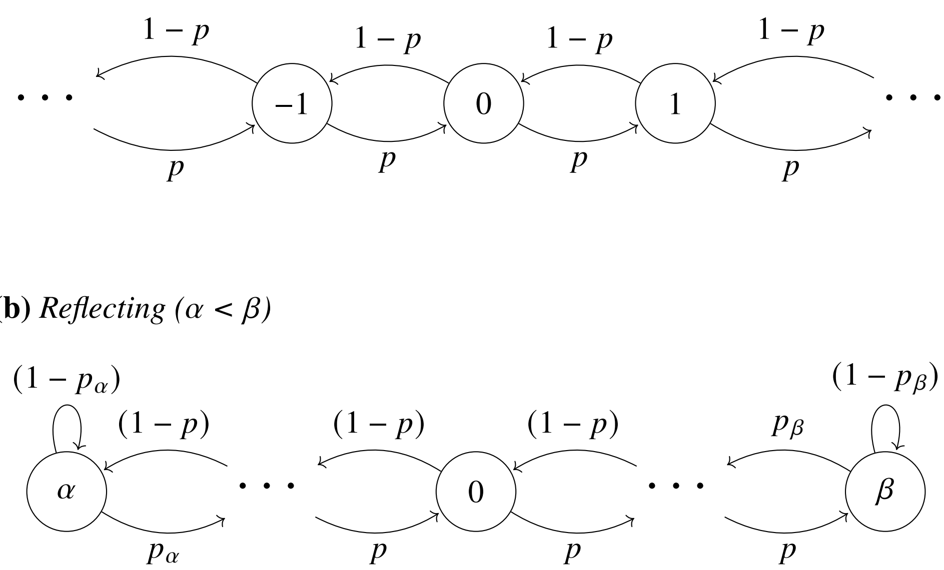
with the convention that $\gcd\{\emptyset\}$ is infinite. If $\gamma_j = 1$, then s_j is *aperiodic*. Proving that periodicity is a class property involves some tedious number theory, but Levin et al. (2008) presents the standard proof. Recall that irreducibility only requires that, for each pair of states, there is a time t such that a t-step transition between them may occur. A standard result of aperiodicity is the following: For all pairs of states, there is a time t' such that for this time and all

Figure 2

Random Walks on $\mathbb{Z}$ ($\mathcal{X}_0 = 0$)

(a) *No Boundaries*

(b) *Reflecting ($\alpha < \beta$)*



times thereafter, a transition from one to the other may occur. An aperiodic chain ensures there is a time past which any state may be visited from any other state.

Recurrence

It is natural to ask how often and at what rate individual states are revisited inside a communicating class. The *total number of visits to state s_j* is the random variable

$$\mathcal{V}_j := \sum_{t=0}^{\infty} \delta\left(\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_j\right) , \tag{17}$$

where $\delta(\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_j) := 1$ when the event $\{\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_j\}$ occurs

$$\delta(\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_j) := \begin{cases} 1\,, & \text{when the event } \left\{\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_j\right\} \text{ occurs;} \\ 0\,, & \text{otherwise.} \end{cases} \tag{18}$$

For $n \geq 0$, the n^{th} *return time of state* s_j is

$$\mathcal{T}_j^{(n)} := \begin{cases} \min\left\{t > \mathcal{T}_j^{(n-1)} \,\middle|\, \mathcal{X}_t = s_j,\ \mathcal{X}_0 = s_j\right\}, & \text{when } n > 0\,; \\ 0\,, & \text{when } n = 0\,. \end{cases} \tag{19}$$

By convention, the quantity $\min\{\emptyset\}$ is infinite. Note that, for all $k \geq 0$,

$$\mathcal{T}_j^{(k)} > \mathcal{T}_j^{(k-1)} > 0. \tag{20}$$

Analyzing Return Times

The Markov property was stated in terms of a fixed time t, but rather than fixing the time when a state is visited, we could instead state the Markov property in terms of a *stopping time* T. This is a possibly infinite random variable such that, for some time m, the states up to time m completely determine the stopping event $\{T = m\}$; no knowledge of the future is required. Return times are stopping times, since for a chain starting in state s_j, we have the event

$$\left\{\mathcal{T}_j^{(1)} = n\right\} := \left\{\mathcal{X}_k = s_j;\ \mathcal{X}_\ell \neq s_j, 1 \leq \ell < k;\ \mathcal{X}_0 = s_j\right\}\ . \tag{21}$$

The chain satisfies the *strong Markov property* if the Markov property holds for the finite stopping time T: The chain at $\mathcal{X}_T$ is the same chain as the one at $\mathcal{X}_0$. It is rather tedious to show that a discrete-time chain must be strongly Markovian, but a continuous-time chain need not be (Shalizi & Kontorovich, 2007). Thus, in our case, the sequence $\left(\mathcal{T}_j^{(n)}\right)$ over n is independent and identically distributed when $\mathcal{T}_j^{(1)}$ is finite with probability 1, We thus adopt the shorthand

$$\mathcal{T}_j^{(1)} := \mathcal{T}_j\ . \tag{22}$$

Now, the probability that s_j is revisited in finite time is

$$\mathbb{P}\left(\mathcal{T}_j < \infty\right) = \sum_{n=0}^{\infty} \mathbb{P}\left(\mathcal{T}_j^{(n)} = n\right)\ . \tag{23}$$

From this fact, we can tie together $\mathcal{V}_j$ and $\mathcal{T}_j$ for a standard but interesting result (Ross, 2019).

Since the return times are independent and identically distributed, we repeatedly apply $\mathcal{T}_j$:

$$\mathbb{P}\left(\mathcal{V}_j = k\right) = \left(\prod_{t=1}^{k} \mathbb{P}\left(\mathcal{T}_j^{(t)} < \infty\right)\right)\left(1 - \mathbb{P}\left(\mathcal{T}_j^{(k+1)} < \infty\right)\right) \tag{24}$$

$$= \left(\mathbb{P}\left(\mathcal{T}_j < \infty\right)\right)^k \left(1 - \mathbb{P}\left(\mathcal{T}_j < \infty\right)\right). \tag{25}$$

The visits $\mathcal{V}_j$ have geometric probability density with $p = \mathbb{P}\left(\mathcal{T}_j < \infty\right)$. Observe that

$$\mathbb{E}\big(\delta(\mathcal{X}_t = s_j)\big) = \left(0 \cdot \mathbb{P}\left(\mathcal{X}_t \neq s_j\right)\right) + \left(1 \cdot \mathbb{P}\left(\mathcal{X}_t = s_j\right)\right) = \mathbb{P}\left(\mathcal{X}_t = s_j\right). \tag{26}$$

Since $\mathcal{V}_j$ is distributed geometrically, this random variable has mean

$$\mathbb{E}(\mathcal{V}_j) = \lim_{t\to\infty} \mathbb{E}\left(\sum_{k=1}^{t} \delta(\mathcal{X}_k = s_j)\right) = \lim_{t\to\infty} \sum_{k=1}^{t} \mathbb{P}\left(\mathcal{X}_k = s_j\right) = \frac{1}{1 - \mathbb{P}\left(\mathcal{T}_j < \infty\right)}. \tag{27}$$

Thus, $\mathbb{E}(\mathcal{V}_j)$ diverges if and only if $\mathbb{P}\left(\mathcal{T}_j < \infty\right)$ approaches 1 from the left. Since we have a probability, we only care about this limit. A state s_j is *recurrent* if $\mathbb{P}\left(\mathcal{T}_j < \infty\right) = 1$ and *transient* if $\mathbb{P}\left(\mathcal{T}_j < \infty\right) < 1$. Intuitively, a recurrent state will be visited infinitely often in asymptotic time, so

$$\mathbb{P}\left(\mathcal{V}_j = k\right) = 0 \tag{28}$$

for all finite k. A transient state will be visited only finitely many times. Table 1 shows how (27) gives multiple ways to understand recurrence.

Just like reducibility, the classifications of recurrence have nice communication properties. Let state s_i be recurrent and communicate with s_j. Thus, there are times t and t' such that $\left(\mathcal{P}^t\right) > 0$ and $\left(\mathcal{P}^{t'}\right) > 0$. Now let t'' be an arbitrary time. By an argument like that in (10),

$$\sum_{t''=0}^{\infty} \left(\mathcal{P}^{t+t'+t''}\right)_{j,j} \geq \left(\mathcal{P}^{t'}\right)_{j,i} \left(\sum_{t''=0}^{\infty} \left(\mathcal{P}^{t''}\right)_{i,i}\right) \left(\mathcal{P}^t\right)_{i,j}. \tag{29}$$

Table 1

Comparison of Conditions for Recurrence and Transience

Recurrence	$\mathbb{P}\left(\mathcal{T}_j < \infty\right) = 1 \iff$	$\mathbb{E}(\mathcal{V}_j)$ infinite $\iff$	$\sum_{t=1}^{\infty} \mathbb{P}\left(\mathcal{X}_t = s_j\right)$ diverges $\implies$	$\lim_{t\to\infty} \mathbb{P}\left(\mathcal{X}_t = s_j\right) > 0$
Transience	$\mathbb{P}\left(\mathcal{T}_j < \infty\right) < 1 \iff$	$\mathbb{E}(\mathcal{V}_j)$ finite $\iff$	$\sum_{t=1}^{\infty} \mathbb{P}\left(\mathcal{X}_t = s_j\right)$ converges $\implies$	$\lim_{t\to\infty} \mathbb{P}\left(\mathcal{X}_t = s_j\right) = 0$

Because s_i is recurrent, the right-hand side diverges. The left-hand side then also diverges by comparison, and so s_j is also recurrent. Hence, recurrence is a class property. An entirely parallel argument shows the same for transience. With these results, an irreducible chain will be referenced as simply a recurrent or transient chain.

Specifying Recurrence

When an arbitrary state s_j is transient, the convergence of the respective sum requires

$$\lim_{t\to\infty} \mathbb{P}\left(\mathcal{X}_t = s_j\right) = 0\,. \tag{30}$$

The converse is not true, because having the terms of a sequence approach zero is necessary but not sufficient for the respective series to converge. So a state may have its probability of visitation vanish but not vanish fast enough; that is, it may still be recurrent. Since we already know that the symmetric simple walk is irreducible, we will show its recurrence by considering an arbitrary state $s_j \in \mathbb{Z}$ with $\mathcal{X}_0 = s_j$. Since steps in this walk are taken one unit at a time, the chain only returns to s through an equal number of moves left and right. Hence, for times $t > 0$,

$$\mathbb{P}\left(\mathcal{X}_t = s_j\right) = \begin{cases} \binom{2t}{t} p^t (1-p)^t\,, & \text{when } 2 \mid t\,; \\ 0\,, & \text{when } 2 \nmid t\,. \end{cases} \tag{31}$$

Here, Levin et al. (2008) present a useful simplification based upon Stirling's approximation of the factorial:

$$\lim_{n\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1\,. \tag{32}$$

That is, the quantity $\left(n^n e^{-n}\sqrt{2\pi n}\right)$ is asymptotically $n!$. The notation $\to$ will denote this relationship. Taking the walk as a series of Bernoulli experiments, the approximation yields

$$0 \leq \mathbb{P}\left(\mathcal{X}_t = s\right) = \binom{2t}{t} p^t (1-p)^t \to \frac{2^{2t}}{\sqrt{t\pi}} p^t (1-p)^t = \frac{\left(4p(1-p)\right)^t}{\sqrt{t\pi}} \leq \left(4p(1-p)\right)^t\,. \tag{33}$$

When $p \neq \frac{1}{2}$ (the non-symmetric case), we have $4p(1-p) < 1$. So

$$\sum_{t=0}^{\infty} \mathbb{P}\left(\mathcal{X}_t = s\right) \leq \sum_{t=0}^{\infty} 4p(1-p)^t = \frac{1}{1-4p(1-p)}\,. \tag{34}$$

Hence the state 0 will be visited a finite number of times, so the non-symmetric simple random walk is transient. As Figure 3a shows, such transience becomes evident with even a $\frac{1}{2}$% asymmetry in p: The example walks wander off toward negative infinity and on average will never be positive.

In the symmetric case, however, we have $4p(1-p) = 1$. Hence,

$$0 \leq \mathbb{P}\left(\mathcal{X}_t = 0\right) = \binom{2t}{t}\left(\frac{1}{2}\right)^{2t} \to \frac{1}{\sqrt{t\pi}}\,. \tag{35}$$

Then,

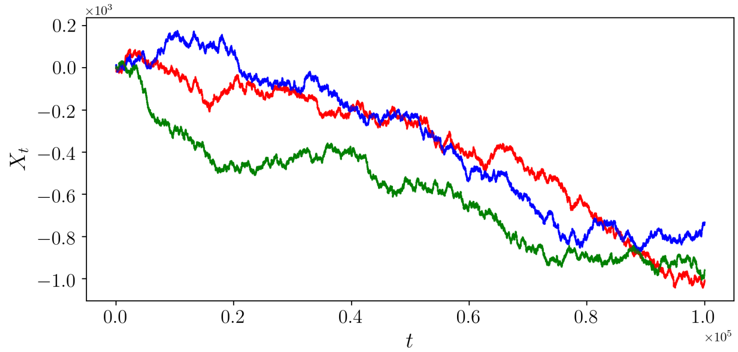
$$\sum_{t=0}^{\infty} \mathbb{P}\left(\mathcal{X}_t = s\right) \geq \sum_{t=0}^{\infty} \frac{1}{t}\,, \tag{36}$$

which is the divergent harmonic series. Hence the state will be visited infinitely often. As Figure 3b illustrates, the symmetric simple random walk is recurrent. In fact, Pólya (1921) also proved

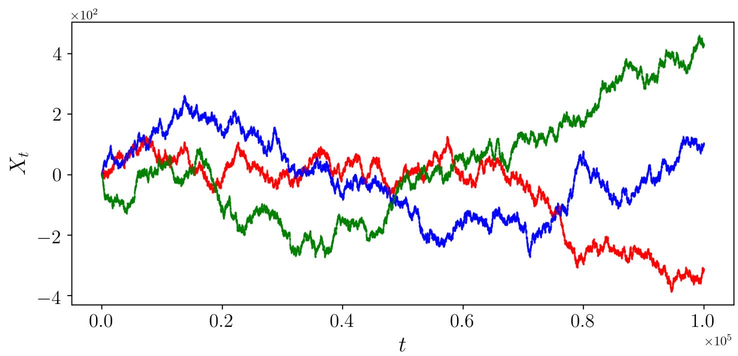
Figure 3

Empirical Illustrations of Simple Random Walks

(a) *Transient Simple Random Walks* ($p = 0.495$; $X_0 = 0$)



(b) *Recurrent Simple Random Walks* ($p = 0.500$; $X_0 = 0$)



that the symmetric random walk on $\mathbb{Z}^2$ is also recurrent but transient on $\mathbb{Z}^3$. In the one-dimensional recurrent case,

$$\lim_{t\to\infty} \mathbb{P}\left(X_t = s\right) \to \lim_{t\to\infty} \frac{1}{\sqrt{t\pi}} = 0\,. \tag{37}$$

This is a state that falls on the edge of recurrence and transience. Refining our definition of recurrence will provoke a deep understanding of a chain's asymptotic behavior.

Recurrence and Generating Functions

We follow Zelen (2005) in presenting a probabality generating function result to more finely define recurrence. For a nonnegative sequence (a_n) of real numbers and $s \in \mathbb{R}$ such that $|s| \leq 1$, a *generating function* is a power series with unit radius of convergence:

$$G(s) := \sum_{n=0}^{\infty} a_n s^n . \tag{38}$$

First, by Abel's theorem for power series, we have the left-hand limit

$$\lim_{s \to 1} G(s) = \sum_{k=0}^{\infty} a_k . \tag{39}$$

Recall from analysis if (b_n) is a sequence of real numbers with

$$\lim_{n \to \infty} b_n = b , \tag{40}$$

then also the *Cesàro sum* also converges:

$$\sum_{n=0}^{\infty} \frac{b_n}{n+1} = b . \tag{41}$$

Hence, when the limit exists, by (41) we have

$$\lim_{n \to \infty} \sum_{k=0}^{n} \frac{a_k}{n+1} = \lim_{n \to \infty} \frac{1}{n+1} \left(\sum_{k=0}^{n} a_k \right) = \lim_{s \to 1} G(s) . \tag{42}$$

Zelen (2005) proves that if a sequence (b_n) has limit b, then the left-hand limit

$$\lim_{s \to 1} \left((1-s) \sum_{n=0}^{\infty} b_n s^n \right) = b. \tag{43}$$

When the limits exist, then, we combine (42) and (43) to obtain

$$\sum_{k=0}^{\infty} \frac{a_k}{k+1} = \lim_{s \to 1} \big((1-s) G(s) \big) . \tag{44}$$

We will soon use this result powerfully.

First, however, apply (21) and the Markov property to obtain the separate result

$$(\mathcal{P}^n)_{j,j} = \sum_{k=1}^{n} \mathbb{P}\left(\mathcal{T}_j = k\right) \cdot \left(\mathcal{P}^{n-k}\right)_{j,j} . \tag{45}$$

For $s \in (0,1)$, define the generating functions

$$T(s) := \sum_{n=0}^{\infty} \mathbb{P}\left(\mathcal{T}_j = n\right) s^n , \tag{46}$$

$$P(s) := \sum_{n=0}^{\infty} (\mathcal{P}^n)_{j,j}\, s^n . \tag{47}$$

As expected from the properties of generating functions, note that by (23),

$$\lim_{s\to 1} T(s) = \sum_{n=1}^{\infty} \mathbb{P}\left(\mathcal{T}_j = n\right) = \mathbb{P}\left(\mathcal{T}_j < \infty\right) , \tag{48}$$

$$\lim_{s\to 1} T'(s) = \sum_{n=1}^{\infty} n\mathbb{P}\left(\mathcal{T}_j = n\right) = \mathbb{E}(\mathcal{T}_j) . \tag{49}$$

We say that a state s_j is *positive recurrent* if $\mathbb{E}(\mathcal{T}_j)$ is finite, and the state is *null recurrent* otherwise. By combining the identity in (45) with the given probability generation functions, we can derive the important identities

$$P(s) - 1 = T(s)P(s) , \tag{50}$$

$$P(s) = \frac{1}{1 - T(s)} . \tag{51}$$

In the symmetric simple random walk example, we follow Mez (2013) to calculate a nice closed-form expression for $P(s)$ from the Taylor expansion of $\mathbb{P}\left(\mathcal{X}_t = s_j\right)$:

$$P(s) = 1 + \sum_{n=1}^{\infty} \binom{2t}{t} p^t (1-p)^t = \frac{1}{\sqrt{1 - 4p(1-p)s^2}}, \text{ when } |s| < \frac{1}{\sqrt{4p(1-p)}} . \tag{52}$$

Then, using the identity of (50),

$$T(s) = 1 - \sqrt{1 - 4p(1-p)s^2}\,. \tag{53}$$

Substituting into (49), and using the fact that $p = \frac{1}{2}$,

$$\lim_{s\to 1} T'(s) = \sum_{n=1}^{\infty} n\mathbb{P}\left(\mathcal{T}_j = n\right) = \lim_{s\to 1} \frac{s}{\sqrt{1-s^2}}\,, \tag{54}$$

which does not exist for $|s| \leq 1$. Thus, $\mathbb{E}(\mathcal{T}_j)$ is infinite, and so the symmetric random walk is null recurrent. It remains to show that positive and null recurrence are also class properties, but showing those facts will take us to the heart of our target result.

Distributions

We can collect the chain's marginal distribution at time t into a row vector $\boldsymbol{\nu}^{(t)} \in \mathbb{R}^{|S|}$ with unit sum. The vector $\boldsymbol{\nu}^{(0)}$ is the *initial distribution* of the chain. Now, given a distribution for $\mathcal{X}_t$,

$$(\boldsymbol{\nu}^{(t+1)})_j = \sum_{s_i \in \mathcal{S}} \left(\mathbb{P}\left(\mathcal{X}_t = s_i\right) \cdot \mathbb{P}\left(\mathcal{X}_{t+1} = s_j \mid \mathcal{X}_t = s_i\right)\right) = \sum_{s_i \in \mathcal{S}} \left((\boldsymbol{\nu}^{(t)})_i \cdot (\mathcal{P})_{i,j}\right) \tag{55}$$

by the Markov property. Figure 4 shows how the marginal distribution for the random web surfer evolves through time for various initializations of p and q. Using Chapman-Kolmogorov, the calculation scales inductively to render a distribution for an arbitrarily future time. In matrix form,

$$\boldsymbol{\nu}^{(t+k)} = \boldsymbol{\nu}^{(t)}\mathcal{P}^k\,. \tag{56}$$

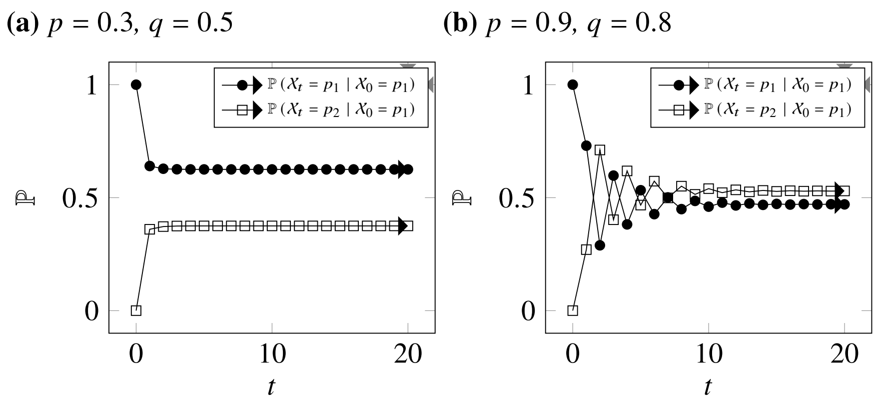
A *stationary distribution* is invariant under $\mathcal{P}$; it is an eigenvector of $\mathcal{P}$ with unit eigenvalue.

For an irreducible chain, define the sequence

$$\left(S_t(j)\right) := \left(\frac{1}{t}\sum_{k=1}^{t} \delta\left(\mathcal{X}_k = s_j\right)\right)\,. \tag{57}$$

Figure 4

Evolving Marginal Distributions for Random Web Surfer



Note that $0 \leq S_t \leq S_{t+1}$ for all times t. When the limits exist, let $\boldsymbol{\rho} \in \mathbb{R}^{|S|}$ be a *vector of limiting proportions* of time that the chain spends in each state s_j, defined elementwise by

$$\rho_j := \lim_{t \to \infty} S_t(j) \,. \tag{58}$$

Note that the event $\{\mathcal{X}_k = s_j\}$ does not depend upon an initial state. Thus, a vector of limiting proportions exists if and only if the component limits exist independent of starting state. When all its components exist, the vector $\boldsymbol{\rho}$ is indeed a probability distribution because the chain by definition visits a state at each timestep, so the long-run proportions of time must sum to unity. Many books jump from the stationary distribution to the limiting distribution discussed below, but we follow Sigman (2009) in treating the vector of limiting proportions first.

Asymptotic Distributions

We now present a more intuitive characterization of the vector of limiting proportions. Recall that if $(\mathcal{X}_n)$ is a monotone sequence of nonnegative random variables, then

$$\mathbb{E}\left(\lim_{n \to \infty} \mathcal{X}_n\right) = \lim_{n \to \infty} \mathbb{E}(\mathcal{X}_n) \,. \tag{59}$$

Wolpert (2018) presents the elementary proof for expectations of Lebesgue's general result – the bounded convergence theorem. Now, let $\boldsymbol{\rho}$ be a vector of limiting proportions, and take the expectation of coordinate ρ_j. By (59) and then (26),

$$\rho_j = \mathbb{E}\left(\lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{1}\delta(\mathcal{X}_k = s_j)\right) = \lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}\mathbb{E}\left(\delta(\mathcal{X}_k = s_j)\right) = \lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}\mathbb{P}\big(\mathcal{X}_k = s_j\big)\,. \tag{60}$$

Now, by (44), we transition to the generating function:

$$\rho_j = \lim_{t\to\infty}\sum_{k=0}^{t}\frac{\mathbb{P}\big(\mathcal{X}_k = s_j\big)}{t+1} = \lim_{s\to 1}\,(1-s)\left(\frac{1}{1-T(s)}\right)\,. \tag{61}$$

When s_j is transient so $\mathbb{P}\left(\mathcal{T}_j < \infty\right) < 1$, the limit vanishes. However, when s_j is recurrent we have the indeterminate form $\frac{0}{0}$. Note that, when the chain has period d,

$$\lim_{s\to 1} T'(s) = \lim_{s\to 1}\frac{\mathrm{d}}{\mathrm{d}s}\left(\sum_{k=0}^{\infty}\mathbb{P}\left(\mathcal{T}_j = k\right)s^{kd}\right) = d\cdot\sum_{k=0}^{\infty}k\mathbb{P}\left(\mathcal{T}_j = k\right) = d\cdot\mathbb{E}(\mathcal{T}_j)\,. \tag{62}$$

Using L'Hôpital's rule to continue from (61),

$$\rho_j = \lim_{s\to 1}\frac{1}{T'(s)} = \frac{d}{\mathbb{E}(\mathcal{T}_j)}\,. \tag{63}$$

When s_i is null recurrent so $\mathbb{E}(\mathcal{T}_j)$ is infinite, the limit also vanishes. Thus, $\boldsymbol{\rho}$ forms a valid probability distribution if and only if all the represented states are positive recurrent. By Chapman-Komolgorov, we may expand the convergence to the entire transition matrix. That is, a vector of limiting proportions exists if and only if

$$\lim_{t\to\infty}\frac{1}{t}\left(\mathcal{P}^t\right) = \begin{bmatrix}\rho_0 & \rho_1 & \cdots\\ \rho_0 & \rho_1 & \cdots\\ \cdots & \cdots & \ddots\end{bmatrix} = \begin{bmatrix}\boldsymbol{\rho}\\ \boldsymbol{\rho}\\ \vdots\end{bmatrix}\,. \tag{64}$$

Now, consider another distribution ρ' that is also a stationary distribution. Then,

$$\boldsymbol{\rho}' = \boldsymbol{\rho}'\mathcal{P} = \boldsymbol{\rho}'\left(\lim_{t\to\infty}\frac{1}{t}\left(\mathcal{P}^t\right)\right) = \boldsymbol{\rho}'\begin{bmatrix}\rho_0 & \rho_1 & \cdots\\ \rho_0 & \rho_1 & \cdots\\ \cdots & \cdots & \ddots\end{bmatrix}\,. \tag{65}$$

Considering an arbitrary component, we obtain

$$\rho'_j = \left(\sum_{s_i \in \mathcal{S}} \rho'_i \right) \rho_j = \rho_j \, , \tag{66}$$

since ρ' is also a distribution and so its components must sum to unity. Thus, when it exists, the vector of limiting proportions is the unique stationary distribution for the chain. When the chain starts with initial distribution ρ, the distribution will thus remain unchanged throughout time. We will soon glimpse the nice properties that such stationarity provides. The preceding arguments for the stationarity and uniqueness of the vector of limiting proportions are inuitively compelling, but Karlin and Taylor (2014) present more rigorous proofs.

Impacts of Periodicity

We ultimately seek to analyze the conditions under which the chain evolves an arbitrary starting distribution into a *limiting distribution* $\boldsymbol{\pi}$ such that, independent of initial state,

$$\pi_j := \lim_{t \to \infty} \mathbb{P}\big(\mathcal{X}_t = s_j\big) \, . \tag{67}$$

Note that the element ρ_j contains a Cesàro sum of the sequence $\left\{ \mathbb{P}\left(\mathcal{X}_t = s_j\right) \right\}$. Since (40) implies (41) but the converse is not true, it is not surprising that convergence to a limiting distribution is also strictly stronger than convergence to a vector of limiting proportions. When the limiting distribution exists, the vector of limiting proportions will coincide with it.

When the chain has period $d > 1$, this convergence holds only for the times $d\mathbb{N}$. If we wish to consider all time, we must perform a time average of the probabilities. To see this, note that when $d > 1$, we may choose a subsequence $\{n_k\}$ of times that contain no return times of s_j. Since $(\mathcal{P}^{n_k})_{j,j} = 0$, convergence to $\pi_j > 0$ cannot hold along the subsequence, so the limiting probability of (67) fails to exist. Hence, aperiodicity is strictly stronger than irreducibility. To see an example of such behavior, consider the random Web surfer again but let $p = q = 1$. Then, for $n \geq 1$, the chain has n-step transitions

$$(\mathcal{P}^n) = \begin{cases} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{when } 2 \mid n\,, \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \text{when } 2 \nmid n\,. \end{cases} \tag{68}$$

For example, we have the sequence

$$\Big((\mathcal{P}^n)_{1,1} \Big) = (1, 0, 1, 0, \cdots)\,, \tag{69}$$

which does not converge, but the sequence of arithmetic means

$$\left(\frac{(\mathcal{P}^n)_{1,1}}{n} \right) = \left(\frac{1}{1}, \frac{1}{2}, \frac{2}{3}, \frac{2}{4}, \cdots \right) \tag{70}$$

converges to $\frac{1}{2}$. Overall, we have vector of limiting proportions

$$\boldsymbol{\nu} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \tag{71}$$

The limits of all the non-averaged component sequences like (69) do not exist, and the chain forever depends upon its starting state. It cannot reach a distribution independent of that state, so the limiting distribution does not exist.

Further Examples

Our Web surfer chain is actually the smallest simple random walk with reflecting boundaries – a clearly recurrent chain. For this case, having a stationary distribution $\boldsymbol{\nu}$ requires

$$\nu_0 = (1 - p)\nu_0 + q\nu_1 \implies p\nu_0 = q\nu_1\,; \tag{72}$$

$$\nu_1 = p\nu_0 + (1 - q)\nu_1 \implies q\nu_0 = p\nu_1\,. \tag{73}$$

We have the further constraint $\nu_0 + \nu_1 = 1$. Hence, the stationary distribution is simply

$$\boldsymbol{\nu} = \left(\frac{q}{p+q}, \frac{p}{p+q}\right) . \tag{74}$$

By induction on the transition matrix, this ν can be shown to be the chain's limiting distribution.

For the simple random walk with no boundaries, however, we have for an arbitrary time t,

$$\nu_j = \sum_{i \in \mathbb{Z}} \left(\nu_i \cdot (\mathcal{P})_{i,j}\right) = p\nu_{j+1} + p\nu_{j-1} . \tag{75}$$

Because of the problem's symmetry across $\mathbb{Z}$, we have that $\cdots = \pi_{-1} = \pi_0 = \pi_1 = \cdots$. However, when $\pi_0 > 0$, an infinite sum of this constant cannot converge. When $\pi_0 = 0$, we would not generate a valid distribution. Thus, in either the null recurrent (symmetric) or transient (non-symmetric) case, this chain does not have a stationary distribution.

Now consider the case of a simple random walk with absorbing boundaries. For this reducible chain, we have stationary distributions

$$\boldsymbol{\nu}_{\varphi} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \text{ and} \tag{76}$$

$$\boldsymbol{\nu}_{\psi} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} . \tag{77}$$

If we modify the transition probabilities so the intermediate states are a closed communicating class, we can form a trivial chain $\mathcal{M}$ on each endpoint and a random walk with reflecting boundary on the inner states. Let ν be the stationary distribution of this sub-chain. Then, for $\alpha, \beta \in [0, 1]$, by considering the block transition matrices we have that

$$\begin{bmatrix} \alpha\boldsymbol{\nu}_{\varphi} & \beta\boldsymbol{\nu} & (1-\alpha-\beta)\boldsymbol{\nu}_{\psi} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ & & \mathcal{P}_{\mathcal{M}} & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{78}$$

provides a stationary distribution of the overall chain. Generally speaking, a reducible chain does not have a unique stationary distribution because the constituent stationary distributions may be linearly combined in infinitely many ways.

Ergodicity and the Strong Law of Large Numbers

A Markov chain is a *stochastic dynamical system* (stochastic process) – dynamical, because it has a state (a point in a given state space), and a set of transition functions which evolve the state through time; stochastic, because its state progresses through random choices governed by transition probabilities (M. Brin & Stuck, 2002). Ludwig Boltzmann (1844-1906) noted that in an isolated system with constant energy, one particle's trajectory followed through enough time will fill the entire state space. Such an insight was his *ergodic hypothesis* – a portmanteau of the Greek ἔργον (work) and ὁδός (path). The idea of tracking points through time – ultimately identifying time averages with space averages – proved useful for analyzing complex physical systems. For this, Boltzmann has been widely recognized as the founder of statistical mechanics, of which dynamical system theory is an abstraction. In 1893, the French mathematician and physicist J. Henri Poincaré (1854-1912) argued that "if [a] system has a fixed total energy that restricts its dynamics to bounded subsets of its [state] space, the system will eventually return as closely as you like to any given initial set of molecular positions and velocities" (Levermore, 2001). The deep Poincaré recurrence theorem remained unproven until 1919, when measure theory had sufficiently matured (Antoniou, 2002).

Return again to the idea of stationary sequences. A sequence $\{X_n\}$ is stationary if, for each $j \geq 0$, the shifted sequence $\{X_{j+n}\}$ has the same distribution as the original sequence. That is, all the finite joint distributions $(X_{n_1}, X_{n_2}, \cdots, X_{n_k})$ match the respective shifted joint distributions

$(\mathcal{X}_{j+n_1}, \mathcal{X}_{j+n_2}, \cdots, \mathcal{X}_{j+n_k})$. A stationary sequence $\{\mathcal{X}_n\}$ is *ergodic* if the sequence cannot be expressed as a union of two distinct stationary sequences. Independently and identically distributed sequences and, as we shall soon see, positive recurrent Markov chains are ergodic sequences. The 1931 Birkhoff-Khinchin ergodic theorem extends the form of (1) to all stationary sequences, and thus presents a much stronger statement than the strong law of large numbers (Sokol & Rønn-Nielsen, 2014). Although ergodicity inheres generally in measure-preserving dynamical systems, measure theory falls outside the scope of this thesis. The generating function approach has provided an elementary argument for such fundamental convergence in Markov chains. We will now see how applying the strong law of large numbers itself yields a more intuitive understanding for the special case of positive recurrent chains.

The Fundamental Theorem

Now, we will show (63) again, using the strong law of large numbers as in Sigman (2009). Take an aperiodic recurrent chain and an arbitrary state s_j inside it. For $n \geq 1$, define

$$\mathcal{Y}_j^{(n)} := \mathcal{T}_j^{(n)} - \mathcal{T}_j^{(n-1)} \tag{79}$$

as the n^{th} *interarrival time of* s_j, so

$$\mathcal{T}_j^{(k)} = \sum_{t=1}^{k} \mathcal{Y}_j^{(t)} . \tag{80}$$

By the strong Markov property, the sequence $\left(\mathcal{Y}_j^{(n)}\right)$ is independent and indentically distributed. In the original definition, we took the limiting proportion with respect to the elapsing time t in the chain. Now, however, we pass to the limiting proportion with respect to the number of visits to each state s_j. Since there are exactly n visits to s_j at the time $\mathcal{T}_j{}^{(n)}$,

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \delta(\mathcal{X}_k = s_j) = \lim_{n\to\infty} \frac{n}{\sum_{k=1}^{n} \mathcal{Y}_j{}^{(k)}} = \lim_{n\to\infty} \frac{n}{\mathcal{T}_j{}^{(n)}} \,. \tag{81}$$

Customarily the vector π is used for the quantity of (81) across the state space. We thus have the limit of a reciprocal sample mean, so by the strong law of large numbers,

$$\mathbb{P}\left(\pi_j = \frac{1}{\mathbb{E}\left(\mathcal{T}_j\right)}\right) = 1\,. \tag{82}$$

The state s_j will, with probability 1, be visited every $\mathbb{E}\left(\mathcal{T}_j\right)$ timesteps. In the transient case, the limiting proportions of time spent in each state all vanish in the limit. In the null recurrent case, the states are visited infinitely often through time, but too infrequently to support any probability mass. Thus, the vector of limiting proportions cannot exist in either case. Since the representation of each π_j is unique, the vector of limiting proportions exists if and only if s_j is positive recurrent.

Further Results

To establish that positive and null recurrence are communication class properties, and thus that these finer definitions of recurrence are consistent, Breuer (2007) uses the inequality method already seen in (10) and (29). Thus, in total, all states of a communicating class must have the same period and together be transient, positive recurrent, or null recurrent. In particular, a finite irreducible chain must be positive recurrent. Consider an irreducible chain with null recurrent state s_i, so all states are null recurrent. By definition, then,

$$\sum_{s_j\in\mathcal{S}} \mathbb{P}\left(\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_i\right) = 1\,, \tag{83}$$

$$\lim_{t\to\infty} \mathbb{P}\left(\mathcal{X}_t = s_j \mid \mathcal{X}_0 = s_i\right) = 0\,. \tag{84}$$

Since we only have finitely many states, the transition probabilities cannot vanish while still summing to unity. So only infinite chains can have null recurrent states.

In an unfortunate quirk of mathematical history, positive recurrent chains are not called ergodic chains (Sigman, 2018). Instead, the Markovian label of ergodicity is reserved for chains that converge to the limiting distribution – the non-averaged limiting probability of being in state s_j. As the example of (69), positive recurrence is not sufficient to guarantee such convergence. An aperiodic positive recurrent chain is called an *ergodic chain*. However, we saw in (66) that ergodicity deals with the time averages of the vector of limiting proportions; aperiodicity is not necessary to exploit the ergodic properties of positive recurrent Markov chains.

The Google Matrix

In the 1895 paper "For the Relative Valuation of Tournament Results," Edmund Landau (1877-1938) corrected a fundamental inconsistency in the era's chess ranking system (Vigna, 2019). Landau showed that naïve power iteration would generate unstable rankings and unbreakable ties sensitive to the number of iterations performed. Instead, he realized that "valor derives from beating strong opponents" (as cited in Boccard, 2020, p. 2). Beating a strong opponent – an opponent with a high probability of triumph – should make oneself stronger than beating a weak opponent, but that strong opponent was made strong by beating other strong opponents, and so on. One's strength consists of the weighted strengths of one's past opponents.

In 1976, Pinski and Narin presented a "self-consistent" methodology of bibliometrics, employing the full strength of the Perron-Frobenius theory that had developed since Landau. During the early days of the public Internet, search engines still relied upon classical techniques to rank all the different aspects of a Web page for response to a search query. By 1998, two Stanford students – Sergey Brin and Larry Page – endeavored to bring order to the Web. The students' graduate project, then hosted at `google.stanford.edu`, did not rely on simple heuristics like URL length or date of publication; their program used the link structure of the Web itself to decide a page's relevance to queries. Their application of ergodic theory to the Internet formed the foundation of the Google enterprise: PageRank.

Exploring PageRank

Like before, each vertex of Figure 5 represents a Web page. Each edge represents the existence of at least one hyperlink between pages p_i and p_j. We seek to calculate the importance of a Web page, and we can view a hyperlink to that page as an endorsement from the linking page. Let N_i^- represent the set of p_i's predecessors, and let N_i^+ represent the set of p_i's successors, so $|N_i^+|$ gives the out-degree of p_i. S. Brin and Page (1998) thus defined the score of page p_i:

$$\pi_i := \sum_{p_j \in N_i^-} \frac{\pi_j}{|N_i^+|} \, . \tag{85}$$

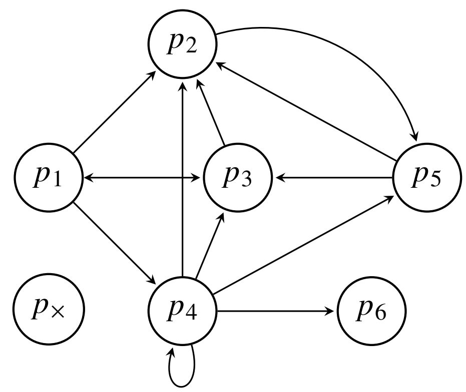
We encode the system thus given in a hyperlink matrix H:

$$(H)_{i,j} := \begin{cases} 1/|N_i^+| \, , & \text{if } p_i \in N_j^- \, ; \\ 0 \, , & \text{otherwise} \, . \end{cases} \tag{86}$$

The hyperlink matrix gives an adjacency matrix weighted to ensure that the probability of transition is 1, as required by the definition of a Markov chain. These marginal distributions need

Figure 5

A Sample Network for Developing the Google Matrix



not be uniform. In the current example, the hyperlink matrix is

$$H = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{87}$$

and this matrix defines a Markov chain over the Web pages. Using the fundamental theorem, we will develop a self-consistent ranking of pages in Figure 5.

First, note that the disconnected node $p_\times$ may be safely removed from our analysis because it has neither inlinks nor outlinks. Page p_6 also has no outlinks; it is a dangling page. However, because it has an inlink, we may not simply discard it. We must connect it to the other pages. Consider a vector $\boldsymbol{\varphi}$, where $\varphi_i := 1$ if page p_i dangles and $\varphi_i := 0$ otherwise. Thus, in our example,

$$\boldsymbol{\varphi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{88}$$

Deriving a limiting distribution from the Web graph still requires more work, for the real Web is quite reducible (Broder et al., 2000). Let $\boldsymbol{\nu}$ be a distribution over all pages in the Web graph. Then, the matrix

$$S := H + \boldsymbol{\varphi}\boldsymbol{\nu}^{\mathsf{T}} \tag{89}$$

gives an irreducible form of Figure 5. The chain given by S is thus positive recurrent. For the H given above, this transformation will put $\frac{1}{6}$ in each entry of the last row.

If an ideal random surfer hits a dangling page, it chooses based upon the distribution $\boldsymbol{\nu}$ what page to next visit. Hence, the Markov chain induced by S is irreducible. However, human Web surfers would often get bored. They will not follow a linear path through the Web; they might suddenly elect to "teleport" – shift course altogether and visit a page on whim. Let $\boldsymbol{\rho}$ be the distribution of teleportation probabilities across the Web, and let T be the rank-one matrix

$$T := \boldsymbol{e}\boldsymbol{\rho}^{\mathsf{T}}, \tag{90}$$

where $\boldsymbol{e}$ is the vector of ones. The original PageRank used a uniform $\boldsymbol{\rho}$, but external factors – a user's search preferences or browsing habits or a search engine's fiat against suspected abusers – can influence the teleportation probabilities. Incorporating a teleportation factor allows the Web chain to become aperiodic. Consider a convex combination of S (irreducible Web structure) and T (teleportation structure) with real *damping factor* $\alpha \in [0, 1]$:

$$G := \alpha S + (1 - \alpha)T = \alpha \left(H + \boldsymbol{\varphi}\boldsymbol{\nu}^{\mathsf{T}}\right) + (1 - \alpha)\left(\boldsymbol{e}\boldsymbol{\rho}^{\mathsf{T}}\right). \tag{91}$$

The matrix G is the *Google matrix*. In Perron-Frobenius terms it is *primitive* and thus has dominant eigenvalue $\lambda_1 = 1$. Since the number of pages on the Web is finite, the induced chain must be positive recurrent. Table 2 presents the example Google matrix for $\alpha = 0.15$, which S. Brin and Page (1998) originally used. We can iterate to approximate the unique limiting

Table 2

Google Matrix for Figure 5 ($\alpha = 0.15$)

$$G \approx \begin{bmatrix} 0.14 & 0.19 & 0.19 & 0.19 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.29 & 0.14 \\ 0.22 & 0.22 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.18 & 0.18 & 0.14 & 0.18 & 0.18 \\ 0.14 & 0.22 & 0.22 & 0.14 & 0.14 & 0.14 \\ 0.17 & 0.17 & 0.17 & 0.17 & 0.17 & 0.17 \end{bmatrix}$$

distribution of the induced chain. As Table 3 shows, the limiting probabilities stabilize quickly. Moreover, a consistent ranking emerges by the second ($k = 2$) iteration:

$$\pi_2 > \pi_5 > \pi_3 > \pi_1 > \pi_4 > \pi_6\,. \tag{92}$$

Interestingly, even though p_4 has five outlinks (the most), it is ranked second-to-last. When the self-link is removed, p_4 actually ranks lower than p_6 by about 2×10^{-3}. The page p_4 ranks so low because its only inlink is p_1, whose only inlink is p_3. In turn p_3's only inlink comes from p_5, but p_5 is linked from the highly-ranked p_2. Thus, the score gets diluted as it moves further from its "source." Now, what if p_4 preserves its self-link and p_2 links back to p_1? At the seventh iteration, we have the ranking $\pi_2 > \pi_1 > \pi_3 > \pi_5 > \pi_4 > \pi_6$, with raw scores

$$\boldsymbol{\pi} = \begin{bmatrix} 0.172046 & 0.183917 & 0.171086 & 0.158787 & 0.163979 & 0.150185 \end{bmatrix}. \tag{93}$$

When p_2 endorses p_1, the endorsement propagates remarkably, and p_5 drops two ranks. Thus, we have explored the impredicative (recursive) nature of the ranking method. To calculate the Google matrix in production, the *power method* of matrix iteration is used (Langville & Meyer, 2006).

The damping factor α controls the fidelity of the model to the raw Web structure. As α approaches 1, the Google matrix becomes more reducible and the power method converges ever

Table 3

Score Iteration for Example Google Matrix

k	π_1	π_2	π_3	π_4	π_5	π_6
0	1	0	0	0	0	0
1	0.141667	0.191667	0.191667	0.191667	0.141667	0.141667
2	0.159583	0.183042	0.168667	0.158042	0.179708	0.150958
3	0.158091	0.184289	0.171639	0.158161	0.177638	0.150182
4	0.158294	0.184266	0.171393	0.158071	0.177809	0.150166
5	0.158275	0.184268	0.171413	0.158078	0.177803	0.150163

more slowly. Moreover, as α approaches 1, the limiting distribution becomes far more sensitive – both in terms of importance score and the consequent PageRanks – to link changes even within minor clusters. With $\alpha = 0.15$, Google's formulation of the power method converged within a thousandth at approximately 50 iterations (S. Brin & Page, 1998).

Conclusion

The law of large numbers enjoys a widely varied application upon stationary sequences of independent and identically distributed random variables. In this thesis, we have motivated basic Markov chain theory and developed a law of large numbers for positive recurrent Markov chains. In so doing, we also developed an elementary background for future studies in ergodicity. Finally, we saw how the stationary distribution of an additionally aperiodic chain provides a powerful way to consistently rank competing objects and undergird the world's most powerful search engine.

References

Antoniou, I. E. (2002, April). Caratheodory and the foundations of thermodynamics and statistical physics. *Foundations of Physics*, *32*(4), 627–641. (https://doi.org/10.1023/A:1015040501205)

Boccard, N. (2020, November). *Landau ranking* (SSRN Scholarly Paper No. ID 2945389). Rochester, NY: Social Science Research Network. (https://papers.ssrn.com/abstract=2945389)

Breuer, L. (2007). *Introduction to stochastic processes*. Canterbury, UK: University of Kent.

Brin, M., & Stuck, G. (2002). *Introduction to dynamical systems*. Cambridge University Press.

Brin, S., & Page, L. (1998, January). *The PageRank citation ranking: Bringing order to the Web* (Tech. Rep.). (http://ilpubs.stanford.edu:8090/422/)

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., . . . Wiener, J. (2000, June). Graph structure in the Web. *Computer Networks*, *33*(1), 309–320. doi: 10.1016/S1389-1286(00)00083-9

Derman, C. (1955). Some contributions to the theory of denumerable Markov chains. *Transactions of the American Mathematical Society*, *79*(2), 541–555. doi: 10.2307/1993046

Dynkin, E. B. (1989). Kolmogorov and the theory of Markov processes. *The Annals of Probability*, *17*(3), 822–832. (https://www.jstor.org/stable/2244385)

Fristedt, B., & Gray, L. (1997). Stationary sequences. In B. Fristedt & L. Gray (Eds.), *A Modern Approach to Probability Theory* (pp. 553–578). Boston, MA: Birkhäuser. (https://doi.org/10.1007/978-1-4899-2837-5_28)

Fujita, Y., Kichikawa, Y., Fujiwara, Y., Souma, W., & Iyetomi, H. (2019, December). Local bow-tie structure of the web. *Applied Network Science*, *4*(1), 1–15. doi: 10.1007/s41109-019-0127-2

Haveliwala, T. H., & Kamvar, S. D. (2003). The Second Eigenvalue of the Google Matrix. , 8.

Karlin, S., & Taylor, H. E. (2014). *A first course in stochastic processes*. Saint Louis: Elsevier Science. (http://qut.eblib.com.au/patron/FullRecord.aspx?p=872574)

Langville, A. N., & Meyer, C. D. (2006). *Google's PageRank and beyond: The science of search engine rankings*. Princeton, N.J: Princeton University Press.

Levermore, D. (2001, August). *Poincare recurrence theorem* [University of Maryland, College Park]. http://www.terpconnect.umd.edu/~lvrmr/History/Recurrence.html.

Levin, D. A., Peres, Y., & Wilmer, E. L. (2008). *Markov chains and mixing times*. American Mathematical Society. (https://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf)

Markov, A. A. (1951). *Izbrannyye trudy*. Leningrad: A.N.S.S.S.R.

Meusel, R., Vigna, S., Lehmberg, O., & Bizer, C. (2014, April). Graph structure in the web — revisited: A trick of the heavy tail. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 427–432). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2567948.2576928

Mez, U. (2013, January). *Null-recurrence of a random walk.* https://math.stackexchange.com/questions/285377/null-recurrence-of-a-random-walk.

Pinski, G., & Narin, F. (1976, January). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management*, *12*(5), 297–312. doi: 10.1016/0306-4573(76)90048-0

Pólya, G. (1921). Uber eine aufgabe betreffend die irrfahrt im strassennetz. *Mathematische Annalen*, *84*, 149–160.

Ross, S. M. (2019). *Introduction to probability models* (Twelfth ed.; S. M. Ross, Ed.). Academic Press (Elsevier). (http://www.sciencedirect.com/science/article/pii/B9780128143469000093)

Schneider, H. (1977, January). The concepts of irreducibility and full indecomposability of a matrix in the works of Frobenius, König and Markov. *Linear Algebra and its Applications*, *18*(2), 139–162. doi: 10.1016/0024-3795(77)90070-2

Seneta, E. (1996). Markov and the birth of chain dependence theory. *International Statistical Review / Revue Internationale de Statistique*, *64*(3), 255–263. doi: 10.2307/1403785

Seneta, E. (2006). Markov and the creation of Markov chains. In *MAM2006: Markov Anniversary Meeting* (p. 20). Charleston, SouthCarolina. (https://www.maths.usyd.edu.au/u/eseneta/senetamcfinal.pdf)

Seneta, E. (2013, September). A tricentenary history of the law of large numbers. *Bernoulli*, *19*(4), 1088–1121. (http://projecteuclid.org/euclid.bj/1377612845)

Shalizi, C. R., & Kontorovich, A. (2007). *Almost none of the theory of stochastic processes*. Duke University. (http://www.stat.cmu.edu/~cshalizi/754/notes/lecture-13.pdf)

Sigman, K. (2009). *Discrete-time Markov chains* [Lecture]. http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-MCI.pdf. Columbia University.

Sigman, K. (2018, August). *Limiting distribution for a Markov chain* [Lecture]. http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-MCII.pdf. Columbia University.

Sokol, A., & Rønn-Nielsen, A. (2014). *Advanced probability*. Copenhagen: Department of Mathematical Sciences, University of Copenhagen.

Vigna, S. (2019, February). Spectral Ranking. *arXiv:0912.0238 [physics]*. (http://arxiv.org/abs/0912.0238)

Wolpert, R. L. (2018, August). Expectation and the Lebesgue theorems. In *STA 711: Probability & Measure Theory* (p. 9). Duke Univeristy. (https://www2.stat.duke.edu/courses/Fall18/sta711/lec/wk-04.pdf)

Zelen, M. (2005, May). *Recurrent and transient States* [Lecture]. http://www.pitt.edu/~super7/19011-20001/19561.pdf. University of Pittsburgh.