The Convergence of Markov Chains and Language

By

Thomas L. LaGrange

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTERS IN SCIENCE

(MATHEMATICS)

at the

CALIFORNIA STATE UNIVERSITY - CHANNEL ISLANDS

2020

APPROVED FOR THE MATHEMATICS PROGRAM

Dec 6, 2019

Dr. Jorge Garcia, Advisor | Date

Dec. 6, 2019

Dr. Roger Roybal | Date

APPROVED FOR THE UNIVERSITY

12/11/19

Dr. Osman Özturgut | Date

DEDICATION

To Kelley. I hope you don't mind that I put down in words…

Acknowledgements

I would like to thank my advisor, Dr. Jorge Garcia, whose encouragement and professionalism guided me through this process. I would also like to thank Dr. Roger Roybal for always being available to help with the many questions that arose during this process.

Camarillo, California

January 2, 2020

Abstract

Markov Chains are a category of stochastic processes with an associated dependence structure. Their inception resulted from A. A. Markov's desire to disprove a rival's assertion concerning the application of the Law of Large Numbers (LLN) to dependent variables. Markov's use of a classic Russian novel to illustrate a dependent relationship between vowels and consonants serves as motivation for investigating the extent to which the properties of these processes can be applied to other means of communication. This work summarizes a history of the Law of Large Numbers and the Markovian properties associated with three different aspects of communication.

TABLE OF CONTENTS

CHAPTER 1

HISTORICAL CONTEXT

1.1 Introduction

In the fall of 2017 we began our investigation into the application of Markovian properties to language. Our original intent was to model the path of a conversation between two or more people as it transitioned from its beginning, through various topics, to its ultimate conclusion. We collected our data by passive observation. We sat in a room that was populated by our coworkers and diagramed the path that their conversations traveled. We made no attempt to participate in or eavesdrop on any conversation. In fact, we excluded any conversation in which we participated and we diagramed only those conversations that were of sufficient volume to be considered for public consumption. As we sorted through our data, we found that though we were able to construct a Markov chain, it was chaotic at best because the state-space (the individual topics of conversation) contained more topics than could be logically considered. As we consolidated and refined our data a pattern began to emerge.

We would normally expect the conversations that take place in a professional environment to be of a professional or otherwise productive nature. As we consolidated the various topics that populated our state-space, we found the opposite to be true. The majority of the conversations in our initial data set typically transitioned into an area that we labeled as unproductive and we hoped to develop some conjecture that would explain this pattern. Unfortunately, our only contribution in this area would be identifying the utility of using Markov chains to map the conversation culture of a given environment. If we wished to develop a deeper insight we needed a better understanding of Markovian properties. Consequently, we felt that this was best accomplished by learning the history of Markov chains and their inception. This turned our research in a slightly different direction and ultimately modified our objectives. We became less concerned with the utility of mapping a conversation culture and more focused on a simple question: Is language Markovian?

To help us answer our new question we chose to investigate language at three different levels. First, we considered it at a basic level of vowels and consonants. We then examined language at a slightly broader level of the individual letters of a given body of work. Finally, we expanded upon our initial investigation into what we call the conversation chain. Our presentation is structured as follows.

In chapter 1, we give a brief history of the development of The Law of Large Numbers and the feud that was the driving force in the development of Markov chains. In chapter 2, we review the structure of Markov chains and the properties which are most applicable to answering our research question. In chapter 3, we present our research into the three levels of language and the applicability of Markovian properties to them. In chapter 4, we present our conclusions, and discuss further research that could be achieved based on the work presented here.

1.2 The Law of Large Numbers

Stochastic processes are collections of random variables indexed by a set (such as the natural numbers) that evolve with time. Markov chains are a category of these processes with a dependence structure as a distinguishing characteristic. This dependence structure is induced by the property that the probability of a chain advancing from one value to a future value (known as states) is dependent only upon the chain's current value or state. Though today we find Markov chains used in areas such as agriculture, economics, engineering, and marketing, their inception resulted from Andrei Markov's desire to disprove a rival's assertion concerning the application of the Law of Large Numbers (LLN) to random variables.

The random variables used in stochastic processes differ from algebraic variables. Regardless of the number of times one attempts to solve for an algebraic variable, the result remains necessarily unchanged; whereas a random variable can assume a number of distinct values in each successive trial. These cases were initially explored in the context of predicting favorable outcomes in games of chance. Consequently specific attention was given to *a priori* computations which were purely deductive in nature. For example, given an urn containing p number of red balls and q number of black balls, the chance of drawing a red ball from the urn, computed $\frac{p}{(p+q)}$, was regarded as a straight forward computation. However, if the numbers p and q were not known and the only available information was "the observations made", then the problem became quite interesting. This problem remained largely unexplored until Jacob Bernoulli addressed this question in 1713 [12].

Figure 1.1: Jacob Bernoulli (1654-1705)

In *Ars Conjectandi*, published eight years after his death, Bernoulli proposed that the probability of a desired outcome could be determined *a posteriori*, meaning that it could be expected that an event will or will not occur as many times in the future as it has been observed to occur in the past under similar circumstances [12]. For example, given an urn containing 3,000 white pebbles and 2,000 black pebbles, by removing and recording the color of one pebble (with replacement) at a time, with a large enough number of draws one could approximate the ratio of white pebbles to black pebbles to be 3:2. Succinctly stated: as the number of trials increases, the closer a proportion will tend towards its theoretical or expected ratio. Hence, the convergence of independent random variables could help approximate such a proportion empirically. Further development of this result continued enroute to Markov's involvement almost 200 years later. We continue with a brief summary of its journey from western Europe into nineteenth century Russia as the contributions of prominent mathematicians extended Bernoulli's Theorem making it a cornerstone of probability theory and the development of the Markov chain's machinery.

The example of choosing pebbles from an urn provides an observer with only two options. The observer could draw a white pebble or a black pebble in each successive trial. These trials, later named Bernoulli Trials, were the impetus behind Bernoulli's Theorem. By defining a favorable result in a trial as a "fertile case", Bernoulli proposed the following:

Let: x be the number of observed fertile cases out of N observations.

Let: p be the unknown proportion of fertile cases.

Then, for any small positive ϵ, and any large c, N may be specified so that

$$P\left(\left|\frac{X}{N} - p\right| \geq \epsilon\right) \leq \frac{1}{(c+1)}$$

This signifies that the probability that the difference between $\frac{X}{N}$ and p is more than ϵ is quite small (less that $\frac{1}{(c+1)}$) if N is large enough. The second implication of this inequality is that by specifying the probability, one could then bound the number of observations required for the fertile cases to fall within the specified probability [12].

In 1733 French mathematician Abraham De Moivre made a major step towards refinement of Bernoulli's Theorem with the publication of the normal approximation to the binomial distribution. First, De Moivre slightly modified Bernoulli's approach of specifying the probability and determining the required N. Instead, he proposed that given some N, one should determine the probability [12]. Secondly, to this point in the life of Bernoulli's Theorem, counting the number of fertile cases when N is very large proved problematic. In his paper, De Moivre specifically noted that adding terms of the binomial with very high powers of n was so difficult that few mathematicians beyond Jacob and Nicholas Bernoulli had ever attempted it [12]. De Moivre's contribution was to approximate the sums of individual Bernoulli Trials (the collection of which become the binomial distribution) when N is very large (as well as approximate) and the probability of a fertile case in a single trial is the symmetric case of $p = \frac{1}{2}$ [11]. In other words, this result provided a method of estimating the number of fertile cases in a series of Bernoulli Trials when the probability of fertile and unfertile cases is equal. In 1812 Pierre Simon Laplace refined De Moivre's result to include those cases where $p \neq \frac{1}{2}$. The combined De Moivre-Laplace theorem would eventually be known as the Central Limit Theorem so named by George Polya in 1920 [3].

In 1837, French mathematician Siméon Denis Poisson's treatise *Recherches sur la probabilité* used the approximation of sums to generalize Bernoulli's Theorem. Poisson's theorem was different from Bernoulli's in that the probability of a fertile event is not assumed to be the same for each trial. Poisson's approach allowed for each trial to have its own particular probability. With this pronouncement, Poisson modified Bernoulli's Theorem by changing the requirement of a common constant probability to the arithmetic mean of the probabilities [9]. Poisson referred to his 1837 contribution as "Loi des grands nombres" or Law of Large Numbers. Perhaps most relevant to our research is the roll that Poisson's contribution played in the Law of Large Number's journey from Europe into Russia. Specifically, Poisson died in 1840 without providing a rigorous proof of his theorem. The first of two such proofs came five years later from P.L. Chebyshev in Russia [11][9].

Chebyshev's 1845 master's thesis proposed an analytic approach to proving Poisson's Law of Large Numbers (LLN) though it went unnoticed at the time and would not be published until after his death. In 1866 however, he delivered a more simplistic proof known as the Chebyshev Inequality that yielded a more general theorem than Poisson's LLN [9] and covered the case previously treated by Bernoulli [2].

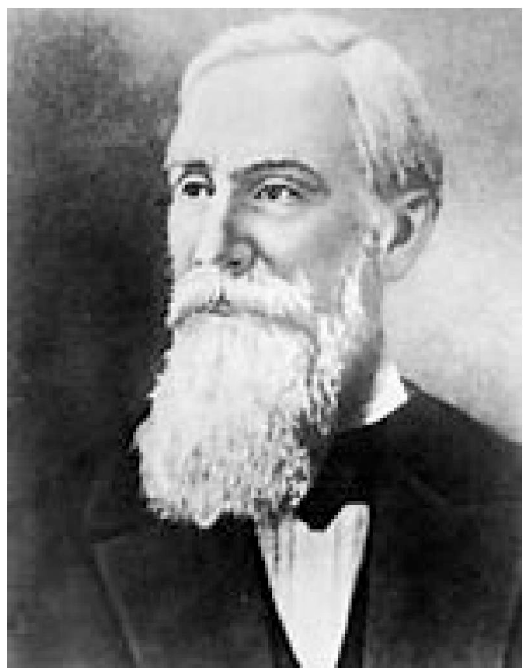


Figure 1.2: P.L. Chebyshev (1821-1894)

1.3 The Fued

Figure 1.3: Pavel Nekrasov (1853-1924)

In 1902, the Moscow Mathematical "School" was steeped in Judeo-Christian ideology. A fundamental tenet of this ideology was the concept of free will or the belief that every person was responsible for their own actions under the guidance of divine providence. Pavel Nekrasov, a leading figure in the Moscow School, held true to this doctrine in his paper entitled *The Philosophy and Logic of Science of Mass Phenomena in Human Activity* (1902) in which he asserted that "independence is a necessary condition for the Law of Large Numbers." [1] Nekrasov argued that free will could not be reduced to a statistical formula and hence the behavior of human beings could not be predicted. In this era, because mathematicians from the Moscow School were loyal supporters of the Russian Orthodox faith, the ideological lens through which they viewed their work diminished their achievements in the eyes of mathematicians from outside their sphere of influence [10]. This case was no exception.

Figure 1.4: A.A. Markov (1856-1922)

Standing in stark contrast to the Moscow School was the secular St. Petersburg Mathematical School influenced heavily by P. L. Chebyshev who had died in 1894. By 1902 Andrei Markov, who was a student and close friend of Chebyshev, had risen to a position of prominence in the St. Petersburg school. Markov also had a direct contribution to the LLN with his inequality (Markov Inequality) that provided a concise proof of the Chebyshev Inequality. Pavel Nekrasov's assertion motivated him to spend several years working to disprove what he called "an abuse of mathematics" [9]. In 1907, he began publishing his research that not only disproved Nekrasov's claim but proffered the concept of chain dependence as well [1]. Of particular importance to our research is Markov's 1913 illustration of chain dependence, based upon his 1907 and 1911 papers, in a lecture to the Royal Academy of Science in St. Petersburg, Russia [8]. For this presentation, Markov extracted the first 20,000 letters from a famous Russian novel, Eugine Onegin, which he separated into either vowels or consonants. Using this corpus, he then identified a dependent relationship between pairs of letters based on the presence and frequency of four factors- vowels following vowels, consonants following vowels, consonants following consonants, and vowels following consonants. His result here reinforced his 1907 conclusion that "independence of quantities does not constitute a necessary condition for the existence of the law of large numbers" [10] and thus extended Bernoulli's result to dependent random variables. Since language, whether spoken or written, is exclusively the product of human thought and action one can conclude that language constitutes human behavior. In disproving Nekrasov's claim that human behavior could not be predicted, Markov showed that the relationship between vowels and consonants within his 20,000-letter corpus formed a structure that converged to values that could be predicted and therefore the LLN could apply to dependent variables. This is important to our research because we will use the notion of convergence to evaluate our question of whether or not language is Markovian. A more contemporary version of Markov's Eugine Onegin illustration will be presented here in due course. It should also be noted that Markov's efforts constituted an early and momentous effort to define language in terms of mathematics [7].

CHAPTER 2

MARKOV CHAINS

2.1 The Markov Property

As stated, A Markov Chain is a system of random variables indexed by some finite or countable classifying set such as the natural numbers. These indexed random variables create a discrete-time state-space, φ, defined by the probabilities associated with transitioning from one state to a another under the restriction that the probability, P, of transitioning to a future state, x_{n+1}, is dependent only upon the chain's current state, x_n. Accordingly, the Markov Property is defined as follows:

$$P\big(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n\big) = P\big(X_{n+1} = x_{n+1} | X_n = x_n\big)[6]$$

Using the Markov Property, we more precisely define the Markov chains used in this work as a system (X_n) that satisfies the following:

i. $X_1, X_2, ..., X_n, \ldots$ are independent random variables.

ii. (X_n) satisfies the Markov Property

iii. Transition probabilities are stationary (i.e. probabilities are not impacted by n)

The simplest case of a Markov chain is a two-state chain illustrated below:

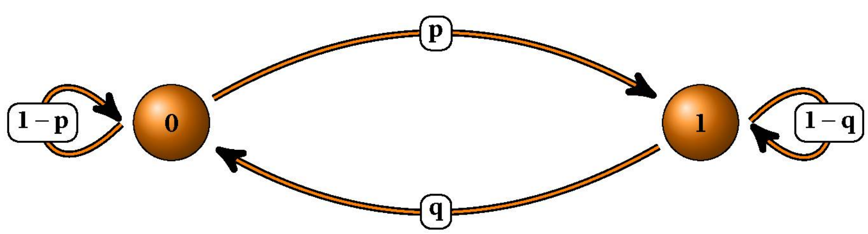
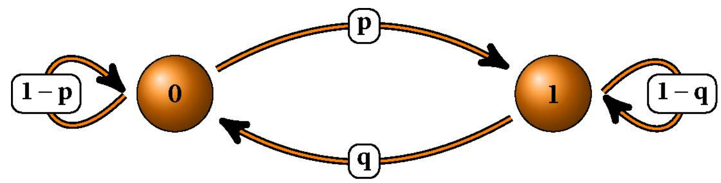


Figure 2.1: Two-state Markov chain



From this example we can see that:

i. The probability of moving from state 0 to state 1 is p, written $P(0,1)$: then,

ii. $P(1,0) = q$

iii. $P(0,0) = 1 - p$ (the complement of p)

iv. $P(1,1) = 1 - q$ (the complement of q)

- $p, q, 1-p, 1-q$ are transition probabilities
- Items i-iv are one-step transition functions which are more generally defined as:

  $P(x,y) = P\big(X_1 = y | X_0 = x\big), x, y \in \varphi$; where

  1. $P(x,y) \geq 0$
  2. $\sum_y P(x,y) = 1$

Using these definitions, we can define chains of larger states, for example:

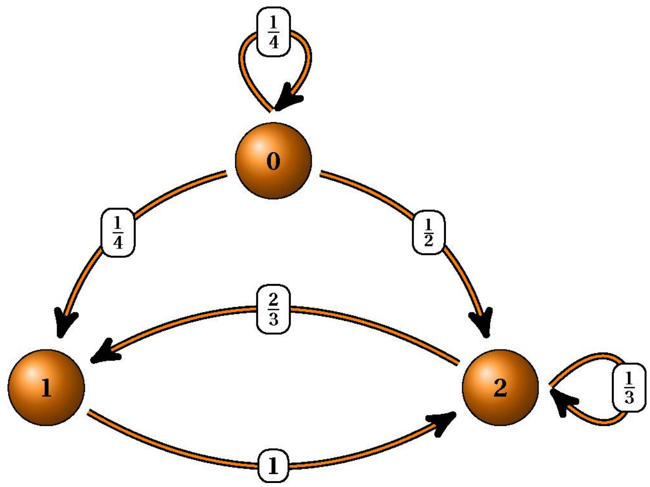


Figure 2.2: Three-state Markov chain

By extending the two-state example, a three-state Markov chain could be defined as:

$$P(0,0) = \tfrac{1}{4} \qquad P(0,1) = \tfrac{1}{4} \qquad P(0,2) = \tfrac{1}{2}$$

$$P(1,0) = 0 \qquad P(1,1) = 0 \qquad P(1,2) = 1$$

$$P(2,0) = 0 \qquad P(2,1) = \tfrac{2}{3} \qquad P(2,2) = \tfrac{1}{3}$$

A more useful tool for delineating the transition probabilities of a Markov chain is the **transition matrix**, P. It contains non-negative elements, with each row summing to 1, that denote the probability of transitioning from one state to another.

$$P = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 1 \\ 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

2.2 Definitions and Examples

The individual states in a Markov chain can be classified into three classes- *recurrent, transient,* and *absorbing.* Let X_n, $n \geq 0$, be a Markov chain within a given state-space with the transition function P. The let $p_{xy} = P_x(T_y < \infty)$ where p_{xy} denotes the probability that a Markov chain starting at state x will be in state y at some positive time or step. Define p_{yy} to be the probability that a Markov chain starting at some point y will ever return to y. Then:

- a state is *recurrent* if $p_{yy} = 1$
- a state is *transient* if $p_{yy} < 1$
- an *absorbing* state is a recurrent state where $P_y(T_y = 1) \Rightarrow p(y, y) = 1$.

Hence, once a chain arrives at an absorbing state, the chain remains in that state.

Another important classification relevant to our presentation is *reducibility.*

- A Markov chain is *irreducible* if each state in the chain has a path to "communicate" with every other state. In other words, A Markov chain is irreducible if you can travel from one state to any other state with positive probability. We say that these states are within the same communication class. An irreducible chain has only one communication class.

- A *reducible* Markov chain is any chain that is not irreducible. A reducible chain can have more than one communication class.

Having presented the basic structure of Markov chains, we now consider a few examples of the numerous chains in use today. These examples are not all inclusive and are presented merely to better the reader's understanding of the uses and structure of the different systems.

1. Random Walk [6]

   Consider a state-space such that $\varphi_1, \varphi_2, \ldots$ are independent integer-valued random variables. Let X_0 also be an integer-valued random variable that is independent of all φ_i within the state-space and let $X_n = X_0 + \varphi_1 + \cdots + \varphi_n$. Then X_n is called a *random walk*. Suppose that a "particle" travels a path along the integers according to this type of Markov chain. When the particle is in state x, regardless of how it got there, it transitions to state y with probability $f(y - x)$.

2. Ehrenfest Chain [6]

   This example is used to describe the exchange of heat or gas molecules between two isolated bodies, proposed in 1906 by Dutch physicist Paul Ehrenfest. Suppose we have two boxes, labeled 1 and 2, and some balls labeled $1, 2, 3, \ldots, d$. Initially some of these balls will be in box 1 with the remainder in box 2. An integer is randomly selected $(1, 2, 3, \ldots, d)$ and the associated ball is removed from the box in which it currently resides and placed into the opposite box. This procedure is repeated indefinitely with each trial being independent. Let X_n denote the number of balls in box 1 after the n^{th} trial. Then X_n is a Markov chain on the space $\{1, 2, 3, \ldots, d\}$. If there are x balls in box 1 at time n, then with probability $\frac{x}{d}$ the ball drawn on the $n+1$ trial will be from box 1 and transferred to box 2. In this case there will be $x-1$ balls in box 1 after the $(n+1)$ trial. Similarly, with probability $\frac{d-x}{d}$ the ball drawn on the $(n+1)$ trial will be drawn from box 2 and transferred to box 1, hence there will be $x+1$ balls in box 1 after the $(n+1)$ trial. The transition function can be expressed as:

$$P(x, y) = \begin{cases} \dfrac{x}{d} & y = x - 1 \\ 1 - \dfrac{x}{d} & y = x + 1 \\ 0 & elsewhere \end{cases}$$

Gambler's Ruin Chain [6]

Suppose a gambler begins with an initial amount of money, in dollars, and begins to make a series of one-dollar bets. Assume that the probability of winning a bet is p, and the probability of losing a bet is q or $(1-p)$. Assume also that if the gambler's money ever reaches zero that he is "ruined" and the amount of money he has remains at zero. Let X_n denote the amount of money the gambler has at time n. Then this is a Markov chain on the space $\{0,1,2,\dots\}$ in which zero is an absorbing state, and for $x \geq 1$

$$P(x,y) = \begin{cases} q & y = x-1 \\ p & y = x+1 \\ 0 & elsewhere \end{cases}$$

CHAPTER 3

MAIN RESULTS

3.1 Case I: Convergence of the VC Chain

A Contemporary Illustration of Chain Dependence at the Level of Vowels and Consonants

Our revised research question was – Is language Markovian? Markov's *Eugene Onegin* illustration of chain dependence provided us with the starting point. Based on Markov's early work, it is established that a dependent relationship between vowels and consonants in the written language does exist. Markov's method for providing a counter-example to Nekrasov's claim employed the statistical methods of the era. He utilized arithmetic means, variation, and the coefficient of dispersion [8] to show that the behavior of the dependence structure between vowels and consonants could be predicted. Rather than reproduce his example here, we chose a more contemporary (and smaller) sample to develop and evaluate our own corpus. Additionally, we employed a more modern method, convergence of the transition matrix, to determine whether the convergence of the chain dependence of vowels and consonants in our corpus could be predicted. To do this, we extracted the first paragraph of Dr. Ivona Grzegorczyk's textbook Mathematics and Fine Arts [4] and removed all special characters and punctuation so that only letters and required spaces remained. From it we identified the following information relevant to our task.

Total Characters: 955 **Vowels:** 306 **Consonants:** 494 **Spaces:** 155

> To the reader - Art is a rare and wonderful creation of the soul, which expresses our personal statements, our likes and dislikes, feelings, thoughts, and love. This book is written to engage readers of any age and interest in the art of repeating patterns and to inspire them to use their own creativity and the newly acquired knowledge of geometric transformations in designing their pictures. It is supposed to broaden the usual way of thinking about artistic creations by introducing some mathematical background and complexity into the designing process. I am always amazed that each student's design emerges through personal expression of himself as he selects colors, patterns and the symmetries he follows. There is no mathematical formula for beauty, and there cannot be. But some interesting mathematical rules may create wonderful patterns! My fondest wish for all readers of this book is that you have great fun, and may you have many new artistic inspirations.

Figure 3.1: The first paragraph of Dr. Ivona Grzegorczyk's Mathematics and Fine Arts [4]

The following relationships were also observed:

- a vowel following a vowel (VV): 47
- a consonant following a vowel (VC): 259
- a vowel following a consonant (CV): 257
- a consonant following a consonant (CC): 237*

* Since there were 800 letters in this corpus, there were only 799 pairs of letters in this corpus. To match the number of pairs of letters with the total number of individual letters, we joined the last letter in the corpus "s" with the first letter in the corpus "t" which added one count of a consonant followed by a consonant (CC) to the total pairs of letters. This brought to total count of pairs of letters to 800 as well.

By counting the total number of letters and letter pairs in the corpus, using the properties of classical probability, we could determine the expected proportion of letter pairs. Specifically:

- Total number of vowels: 306 $\Longrightarrow \dfrac{306}{800} \Longrightarrow \text{P(vowel)} = \underline{0.3825}$
- Total number of consonants: 494 $\Longrightarrow \dfrac{494}{800} \Longrightarrow \text{P(consonant)} = \underline{0.6175}$

With these proportions we should expect the total number of letter pairs in our corpus to be as follows:

- $VV = (0.3825 \times 0.3825) \times \overbrace{(800)}^{\text{total pairs}} \approx \mathbf{117.045}$

- $VC = (0.3825 \times 0.6175) \times (800) \approx \mathbf{188.955}$

- $CV = (0.6175 \times 0.3825) \times (800) \approx \mathbf{188.955}$

- $CC = (0.6175 \times 0.6175) \times (800) \approx \mathbf{305.045}$

However, as previously stated, these counts were not close to the actual number of letter pairs in our corpus:

- VV= **47**

- VC= **259**

- CV= **257**

- CC= **237**

By comparing the transition matrix of the classical expectation with the transition matrix of the empirical data we have the following:

Classical Matrix (Theoretical)

	V	C
V	$\frac{117.045}{306}$	$\frac{188.955}{306}$
C	$\frac{188.955}{494}$	$\frac{305.045}{494}$

$$\implies P = \begin{pmatrix} \frac{117.045}{306} & \frac{188.955}{306} \\ \frac{188.955}{494} & \frac{305.045}{494} \end{pmatrix}$$

Empirical Matrix (Observed)

	V	C
V	$\frac{47}{306}$	$\frac{259}{306}$
C	$\frac{257}{494}$	$\frac{237}{494}$

$$\implies P = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix}$$

Notice the denominators in the vowel and consonant rows of each matrix above. As outlined in chapter 2, these matrices define the probability of transitioning from one state to another. For example, and without loss of generality, consider the case of transitioning from a vowel to a vowel. Since there are 306 vowels in our corpus, there are only 306 possible vowel states, hence the chain cannot transition from a vowel to another vowel unless it currently resides in one of the 306 vowel states. Notice also that these chains are irreducible which allows each state to communicate with every other state in the chain.

As we would expect, the proportions in the classical matrix are those of the proportions of vowels and consonants in the corpus itself, and this can be seen by converting the fractions in each matrix box to a decimal.

$$P = \begin{pmatrix} \frac{117.045}{306} & \frac{188.955}{306} \\ \frac{188.955}{494} & \frac{305.045}{494} \end{pmatrix} = \begin{pmatrix} 0.3825 & 0.6175 \\ 0.3825 & 0.6175 \end{pmatrix}$$

In order to evaluate P for convergence, we introduce the following definitions and properties:

(I) $\pi_0(x) = P(X_0 = x)$ is called the *initial distribution* where;

(i) $\pi_0(x) \geq 0$

(i) $\sum_x \pi_0(x) = 1$

(II) The *joint distribution* of the system can be expressed in terms of the initial distribution and the one-step transition function, hence:

$$P(X_0 = x_0,\ X_1 = x_1) = P(X_0 = x_0)\, P(X_1 = x_1 \mid X_0 = x_0)$$

$$= \pi_0(x_0)\, P(x_0, x_1) \qquad \text{then,}$$

$$P(X_2 = x_2 \mid X_0 = x_0,\ X_1 = x_1) = P(X_2 = x_2 \mid X_1 = x_1) = P(X_1 = x_2 \mid X_0 = x_1)$$

$$= P(x_1, x_2)$$

It follows that this property can be extended to show that:

$$P(X_0 = x_0, \ldots, X_n = x_n) = \pi_0(x_0)\, P(x_0, x_1) \ldots P(x_{n-1}, x_n)$$

The above properties provide us with the method for evaluating each individual one-step transition function as the chain advances from the initial distribution through the chain to its ultimate convergence. We illustrate this process in the following way. If we define an initial distribution , $\pi_0(x)$, as a row vector that indicates a starting point of the chain (vowel or consonant within the corpus) we then multiply this row vector by the classical transition matrix to determine the associated transition probabilities at that point in the chain. For example, and without loss of generality:

Let $\pi_0(x) = [\pi_0(\text{vowel}), \pi_0(\text{consonant})] = [1 \;\; 0]$ define the initial state of the chain as a vowel. Then by multiplying the the given row vector by the classical transition matrix we have:

$$[1 \;\; 0] \times \begin{pmatrix} 0.3825 & 0.6175 \\ 0.3825 & 0.6175 \end{pmatrix} = [0.3825 \;\; 0.6175]$$

This indicates that the probability of transitioning from a vowel to a vowel is 0.3825 and the probability of transitioning from a vowel to a consonant is 0.6175. This then provides us with a new distribution from which we can progress through the system.

We then multiple the transition matrix by the resulting row vector (new distribution) which represents a second step in the chain to obtain:

$$[0.3825 \;\; 0.6175] \times \begin{pmatrix} 0.3825 & 0.6175 \\ 0.3825 & 0.6175 \end{pmatrix} = [0.3825 \;\; 0.6175]$$

for illustration purposes, if we substitute $\pi_0(x) = [0\ \ 1]$ into the classical matrix we would also obtain $[0.3825 \quad 0.6175]$.

What this signifies is that as the classical chain progresses, the transition probabilities in the matrix immediately converge to the initial vowel/consonant proportions, also called *stationary distribution* of the corpus,

$$\begin{pmatrix} 0.3825 & 0.6175 \\ 0.3825 & 0.6175 \end{pmatrix}$$

Therefore, regardless of our state in the classical chain (letter in the corpus) the probability of advancing to a vowel or consonant is equal to the stationary distribution of vowels and consonants in the corpus. But what about the empirical transition matrix which does not match the classical transition matrix? Using the same procedures as above we have:

$$P = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \Longrightarrow \begin{pmatrix} 0.1536 & 0.8464 \\ 0.5202 & 0.4798 \end{pmatrix}$$

Let $\pi_0(x) = [1\ 0]$ define the initial state of the chain as a vowel.

Let $\pi_0(x) = [0\ 1]$ define the initial state of the chain as a consonant.

Using the fractional form of the empirical matrix to reduce rounding errors:

Initial state: **Vowel**

$$n=1: \quad [1 \;\; 0] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.1536 \;\; 0.8464]$$

$$n=2: \quad [0.1536 \;\; 0.8464] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.4639 \;\; 0.5361]$$

$$n=3: \quad [0.4639 \;\; 0.5361] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3502 \;\; 0.6498]$$

$$n=4: \quad [0.3502 \;\; 0.6498] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3918 \;\; 0.6082]$$

$$n=5: \quad [0.3918 \;\; 0.6082] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3766 \;\; 0.6234]$$

$$n=6: \quad [0.3766 \;\; 0.6234] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3822 \;\; 0.6178]$$

$$n = 7: \quad [0.3822 \;\; 0.6178] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3801 \;\; 0.6199]$$

$$n = 8: \quad [0.3801 \;\; 0.6199] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3809 \;\; 0.6191]$$

$$n = 9: \quad [0.3809 \;\; 0.6191] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3806 \;\; 0.6194]$$

$$n = 10: \quad [0.3806 \;\; 0.6194] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3807 \;\; 0.6193]$$

Initial state: **Consonant**

$$n = 1: \quad [0 \;\; 1] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.5202 \;\; 0.4798]$$

$$n = 2: \quad [0.5202 \;\; 0.4798] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3295 \;\; 0.6705]$$

$$n = 3: \ [0.3295 \ \ 0.6705] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3994 \ \ 0.6006]$$

$$n = 4: \ [0.3994 \ \ 0.6006] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3738 \ \ 0.6262]$$

$$n = 5: \ [0.3738 \ \ 0.6262] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3832 \ \ 0.6168]$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$n = 9: \ [0.3805 \ \ 0.6195] \times \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx [0.3807 \ \ 0.6193]$$

Hence, regardless of the starting point in the empirical chain, it will converge to the approximate stationary distribution of vowels and consonants in the corpus as well:

$$\begin{pmatrix} 0.3807 & 0.6193 \\ 0.3807 & 0.6193 \end{pmatrix} \approx \begin{pmatrix} 0.3825 & 0.6175 \\ 0.3825 & 0.6175 \end{pmatrix}$$

A more efficient method of determining the convergence of a transition matrix is the method of raising the matrix, P, to a given value of n. In the case of our empirical chain matrix, based on the previous result, we would raise P^n to a value of $n = 10$, specifically:

$$P^1 = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx \begin{pmatrix} 0.1536 & 0.8464 \\ 0.5202 & 0.4798 \end{pmatrix}$$

$$P^2 = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx \begin{pmatrix} 0.4639 & 0.5361 \\ 0.3295 & 0.6705 \end{pmatrix}$$

$$P^3 = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx \begin{pmatrix} 0.3502 & 0.6498 \\ 0.3994 & 0.6006 \end{pmatrix}$$

$$P^4 = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx \begin{pmatrix} 0.3918 & 0.6082 \\ 0.3738 & 0.6262 \end{pmatrix}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$P^{10} = \begin{pmatrix} \frac{47}{306} & \frac{259}{306} \\ \frac{257}{494} & \frac{237}{494} \end{pmatrix} \approx \begin{pmatrix} 0.3807 & 0.6193 \\ 0.3807 & 0.6193 \end{pmatrix} \approx \begin{pmatrix} 0.3825 & 0.6175 \\ 0.3825 & 0.6175 \end{pmatrix}$$

Regardless of which method we utilize, the result is the same. Since the chain defined by the empirical values of the vowel/consonant relationship is irreducible, it is "well-behaved" and therefore predictable as the number of trials or steps in the chain increase.

3.2 Case II: Convergence of the Letter Chain

In Case I we examined the applicability of Markov properties to a basic level of communication-the relationship of vowels and consonants within a corpus. In this case we will consider a slightly broader view of communication (in written form) and examine the dependence structure between the individual letters and spaces of the same corpus. Our objective here is to determine whether or not the transition matrix will converge to predictable values. What is problematic is that once we evaluate the matrix structure, we must then consider the context of the chain and account for the fact that our "unknown quantity" is the content of the corpus itself, hence the term predictable values may not be clearly defined. For example, using empirical probability, we wish to construct a Markov chain whose state-space is defined by the probability of advancing from one letter to another in such a manner that each current and successive step in the chain combine to reconstruct the actual text of the corpus. Our vehicle for this task is called the *Drivel Generator* developed my Brian Hayes [5]. This software randomly generates text utilizing the dependence structure and the associated probabilities of our corpus.

We start by removing the context of the character chain and consider only the dependence structure of the letters and spaces in our corpus. Because we must consider the spaces within the corpus, we added the 155 spaces to the 800 vowels and consonants and constructed a 26 X 26 matrix (the corpus does not contain the letter j) to determine the dependence structure between the individual characters as shown below. We used MATLAB and Microsoft Excel to compile and display our matrix.

	SP	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
SP	0	0.1558	0.0519	0.0455	0.026	0.0325	0.0455	0.013	0.026	0.0974	0.0065	0.013	0.0455	0.0195	0.0649	0.0455	0	0.039	0.0519	0.1558	0.013	0	0.039	0	0.013	0
A	0.0286	0	0.0143	0.0429	0.0571	0	0	0.0286	0	0	0	0.1143	0.0286	0.2	0	0	0	0.0714	0.0143	0.2857	0.0143	0.0286	0	0	0.0571	0.0143
B	0	0.125	0	0	0	0.25	0	0	0	0	0	0	0	0	0.375	0	0	0.125	0	0	0.125	0	0	0	0	0
C	0.1429	0.1905	0	0	0	0.0476	0	0	0.0952	0.0476	0.0476	0	0	0	0.0952	0	0.0476	0.1905	0	0.0952	0	0	0	0	0	0
D	0.5	0	0	0	0	0.3929	0	0.0357	0	0.0357	0	0	0	0	0	0	0	0	0	0	0.0357	0	0	0	0	0
E	0.2887	0.0722	0	0.0103	0.0412	0.0103	0	0	0	0.0206	0	0.0309	0.0619	0.0515	0.0103	0.0103	0	0.1546	0.1649	0.0206	0	0	0.0206	0.0309	0	0
F	0.4444	0	0	0	0	0.0556	0	0	0	0	0	0	0	0	0.3333	0	0	0	0	0	0.1667	0	0	0	0	0
G	0.3	0.05	0	0	0	0.25	0	0	0.1	0	0	0	0	0.15	0	0	0	0.1	0.05	0	0	0	0	0	0	0
H	0.1471	0.1176	0	0	0	0.5	0	0	0	0.1471	0	0	0	0	0.0294	0	0	0.0294	0	0.0294	0	0	0	0	0	0
I	0.0172	0	0	0.1379	0	0.0172	0	0.0517	0	0	0.0345	0	0.0172	0.2759	0.0862	0	0	0.0862	0.1897	0.069	0	0.0172	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0.2857	0	0	0	0	0.2857	0	0.1429	0	0.1429	0	0	0.1429	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0.3846	0.0385	0	0	0	0.1538	0.0385	0	0	0.1154	0	0.0769	0	0	0.1154	0	0	0	0	0	0	0	0.0385	0	0.0385	0
M	0.0833	0.4583	0	0	0	0.25	0	0	0	0	0	0	0.0417	0	0	0.0417	0	0	0.0417	0	0.0417	0	0	0	0.0417	0
N	0.1607	0.0357	0	0	0.25	0.0179	0	0.1429	0	0.0357	0.0179	0	0	0.0179	0.0357	0	0	0	0.1429	0.1071	0	0	0	0	0.0357	0
O	0.1186	0.0169	0	0.0169	0.0169	0	0.1186	0	0	0	0.0339	0.0339	0.0678	0.1695	0.0508	0	0	0.1017	0.0339	0.0169	0.1356	0.0169	0.0508	0	0	0
P	0	0.2	0	0	0	0.2	0	0	0	0.2	0	0.0667	0	0	0.0667	0.0667	0	0.2	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
R	0.1429	0.0357	0	0	0	0.3393	0.0357	0.0179	0	0.0536	0	0	0.0357	0.0536	0.1071	0	0	0	0.0893	0.0714	0.0179	0	0	0	0	0
S	0.5	0	0	0	0	0.0781	0.0156	0	0.0156	0.0625	0	0.0156	0	0	0.0781	0.0313	0	0	0.0469	0.1094	0.0313	0	0	0	0.0156	0
T	0.1392	0.0127	0	0	0	0.1013	0	0	0.2785	0.1772	0	0	0	0	0.0759	0	0	0.0506	0.0506	0.0506	0.0253	0	0	0	0.038	0
U	0.087	0.0435	0	0.0435	0.0435	0	0	0.087	0	0.0435	0	0.2174	0	0.087	0	0.0435	0	0.087	0.087	0.1304	0	0	0	0	0	0
V	0	0	0	0	0	0.6	0	0	0	0.2	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0	0	0
W	0.0833	0.1667	0	0	0	0	0	0	0.0833	0.0833	0	0.1667	0	0.0833	0.1667	0	0	0.0833	0.0833	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0.3333	0	0	0	0	0	0.6667	0	0	0	0	0	0	0	0	0	0
Y	0.7333	0	0	0	0	0	0	0	0	0	0	0	0.0667	0	0	0.1333	0	0	0.0667	0	0	0	0	0	0	0
Z	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3.2: Letter matrix

This chain once again is irreducible. As we raise the power of the transition matrix, P^n, to $n = 11$ we find (as expected) that the chain also converges to its stationary distribution as shown below.

	SP	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
SP	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
A	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
B	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
C	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
D	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
E	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
F	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
G	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
H	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
I	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
J	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
K	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
L	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
M	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
N	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
O	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
P	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
Q	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
R	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
S	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
T	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
U	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
V	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
W	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
X	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
Y	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001
Z	0.168	0.0699	0.0097	0.0222	0.0295	0.1	0.0189	0.0211	0.0335	0.0612	0.0073	0.027	0.0261	0.0599	0.0596	0.0181	0.0011	0.0586	0.0672	0.0818	0.024	0.0041	0.0127	0.0031	0.0146	0.001

Figure 3.3: Convergence of the Letter matrix

As we stated, we considered the Character chain without regard to context. Consequently what we have shown so far is that the structure of the Character chain converges; but, what happens when we take a slightly broader look at this chain and consider the context of each state? We know that the purpose of joining letters together is to form words, and that we form words in an order or pattern to convey a message. Our objective here was to use Brian Hayes' *Drivel Generator* [5] to reproduce the words in the corpus in a manner that conveys the same message as the corpus.

The *Drivel Generator* uses an algorithm to build a transition matrix that defines sequences of k letters and the probabilities of the various letters that can follow each $k-$character sequence of letters. We load our corpus into the *Drivel Generator* and record the results. For reference here is our corpus.

> To the reader - Art is a rare and wonderful creation of the soul, which expresses our personal statements, our likes and dislikes, feelings, thoughts, and love. This book is written to engage readers of any age and interest in the art of repeating patterns and to inspire them to use their own creativity and the newly acquired knowledge of geometric transformations in designing their pictures. It is supposed to broaden the usual way of thinking about artistic creations by introducing some mathematical background and complexity into the designing process. I am always amazed that each student's design emerges through personal expression of himself as he selects colors, patterns and the symmetries he follows. There is no mathematical formula for beauty, and there cannot be. But some interesting mathematical rules may create wonderful patterns! My fondest wish for all readers of this book is that you have great fun, and may you have many new artistic inspirations.

Note: for proper processing we have removed all capitalization and punctuation.

> is thead ts nd expeny agexprthe s isf isontsthisou rertowsode bor theart abostised ead pis r ong iexper f t is tresout pin souse d asthe cthe br m ad bes eatym ts tirtealy pans heom tispad piempinsfeat y tsfunde icond rmerssf pisusofus anaghadepay thead athem bondeousestof bomeanalws tspanolwalll o w cred be an m be sonde che trt tarosu o atigstheruspe foa and beaderoud ind is ind ped is f e pr ris wore isf adeires a ig ws octstotad mestroureadud tinatomestym crestheansisl enous gsofo be atand l the nd te an as tere tssthemm s s t isfeargsinger deabronstis ion alec patrinst e ans inoul t te amal ig f ikn had i r inof nsthe odesthe ul grgr taul ag e fu rennonstteromedem chexprn k the he i f m oly br t m tt abe w avis tankea cr isexiorg abrme isf thertherty onseano tis d aly is oul indual pal al the al adure ighemeathinathisotisoul he stheanymer irmendexpio d n f ags s bre s pealfounted alys mm woowat ctheatinat wis fowave isleathesf onk pindemeaseathis e pinsf t cay in tio breal ind edu

Figure 3.4: First order Markov chain text reproduction (first attempt)

In the above example we noticed that there are very few "words" randomly generated from the chain created by our corpus. We tried a second time.

> ct emato ms t femaze pe insicrgeanstul at ragrt e stthon m mel ous ty bealf tsy c tireand pitio t eiseereadexpiof on indexpexiowor wountin f am mpesom ereathaulony g watthe ather an wal cathe sf pigsireofus cal n at trg wof ande atred sthead d st he ader sty t be t atad tuly tin malios reac age fuigronstheal ol tstsf ol cagrg oul fono bris e orm of ofulymede theat pis ped pated pag mesof alym me rigs tissofosoave hend bofutssfe g atirem onats wanyme bred foutheisons aremal ouly n c be tisis ty boful eanthesoul piouly th cr forisstouly t ade adur wre t owamallwathinty ousserexpl wosthik iol bon wnd hirexistsin f preangnats s detheatheargremesis watis icrioul pes tul f pars osireatherol icredewo tthaveral isthisf iciredexicreasf watexikn nsticrertusthexind amat r hind reamsonderoust oul pantatamay s wof at sug f ful perisis d ansomavensfomm ande toustision k tis ang the atheaty osf

Figure 3.5: First order Markov chain text reproduction (second attempt)

In the second attempt, we notice that not only are very few words being formed, but that the second attempt does not resemble the first attempt. Therefore, we will redefine the dependence structure of our corpus to create an $n-$order Markov chain which will consider the current state of the chain to be a combination of the $n-$previous states.

N-order Markov Property

$$P(X_n = x_n | X_{n-1} = x_{n-1},\ X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m}) \text{ for } n > m$$

Hence, in second-order Markov chain, the probability of advancing top a future state, X_{n+1}, is dependent only upon the current state, $X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}$. For example, suppose the character chain currently resides at the letter "r" in the word "wonderful" from our corpus. A second-order Markov chain would consider the probability of advancing to a future state given that the current state is "er". The third-order Markov chain would consider the probability of advancing to a future state given hat the current state is "der". This pattern can be extended to the case $n - m$ states. If we were to compute the transition matrix for n-order chains, the size of our matrix (26^n) would be prohibitive. This problem is solved by the *Drivel Generator* . As we proceed we will present only one attempt for each order Markov chain because no two attempts will produce identical results.

> upposesignints bacquir broul boula feelf the wons the gre the the sed the groad-estudere and the symmetroaderns and al reacquirations and istish sou hat ead-estish st istins newly art wonal comergesigninkin cal wons is in all rula folove to exitte and thislikestatish entroaderns is of to exproadergess beations intriess ints himsed the rula rula ruled ints somersometritterestical rul ress perge that is of and thational wonal formay folors to beadereatic to broaden of the derge come supposes anden cometriess intrittergess supposed the newly and the used wonal bouty of this is matteress interns ing piratic to to beacquiresignintem themetries-tureadess thatheirationd destis of thin and them to islikesignintriest wons hathe themetroadesign of and designing theirand dernsformula ruled knows histin-temattersout is matin eatin dis anderge rula rula folove newly artic istintematic is of himsed theme trity ing matterges anderge matte nown the new aboulexity am therge usual rula resigningage thouty a

Figure 3.6: Second-order Markov chain text reproduction

We can see in figure 3.6 that words from the corpus are starting to form.

> backgrough pers of thinking about sometric intere and though pers patterns in-terns matic intere is a rare is no matical follows the repeat you have greatical which studen to engage readen that each state wondersonal which exprestings there is no mation of that each studen to that each expresses into theresses the reate wonders patten their own create wonderful patten to that you have the de-signings that each expresting patten to the some matings the follows the and to their own create wonderful patterns and to this am alway creaden to broaden to engage reations am alway creatic interns am alway of and and mathematic transformula follows thoughts of as he reating mathe readerful patten the read-erful patterns mations and and this by in that each expresses into the reatings their own creadents of the designing processions a rare and and to though pers patterns and love greaden to that each express interess introducing the use there and thround interest wish follows their own creaders of t

Figure 3.7: Third-order Markov chain text reproduction

> soul which expression of himself as he selects colors patterns and the newly acquired knowledge of geometric transformations in designing their own creativity and the newly acquired knowledge of geometric transformations in designing their own creations in designing their own creation of the soul which expression of himself as he selects colors patterns my fondest wish for all readers of this book is that you have many new artistic creations in designing their own creativity and the newly acquired knowledge of geometric transformations in designing their own creation of the soul which expression of himself as he selects colors patterns my fondest wish for all readers of thinking about artistic creation of the soul which expresses our personal expression of himself as he selects colors patterns and to inspirations

Figure 3.8: Fifth-order Markov chain text reproduction

> ways amazed that each students designing their own creativity and complexity introducing some interest in the usual way of this book is that you have great fun and the designing their own creations in designing patterns and complexity introducing some interesting patterns and complexity introducing some mathematical background and interest in the readers of any age and to inspirations in designing their own creations by introducing some mathematical formula for beauty and wonderful patterns and complexity introducing some mathematical background and may create wonderful create wonderful patterns and to inspirations in designing their own create wonderful creations in designing their own creations in designing patterns and the art is a rare and the readers of this book is that each students designing their own creations in designing their own creations in designing process i am always amazed that you have great fun and complexity introducing some interesting mathematical formula for bea

Figure 3.9: Seventh-order Markov chain text reproduction

mathematical formula for beauty and the **newly acquired knowledge of geometric transformations in designing their pictures it is supposed to broaden the usual way of thinking about artistic creation of the soul which expresses our personal expression of himself as he selects colors patterns and the newly acquired knowledge of geometric transformations in designing their pictures it is supposed to broaden the usual way of thinking about artistic creation of the soul which expresses our personal expression of himself as he selects colors patterns** and the symmetries he follows there is no mathematical background and complexity into the **designing their pictures it is supposed to broaden the usual way of thinking about artistic creations by**introducing some mathematical background and complexity into the **designing their pictures it is supposed to broaden the usual way of thinking**

Figure 3.10: Ninth-order Markov chain text reproduction

We initially showed that the dependence structure of the Character chain converged to its stationary distribution at the 11^{th} step in the chain. It should be noted that by the 9^{th} step in the chain, all but a few values in the fourth decimal position, had already converged to the stationary distribution. We can see from the *Drivel Generator* examples that as the order of the chain reaches nine, whole words and ideas form our corpus are beginning to appear. Entire sentences are being repeated with no logical order to them and no two attempts are alike at any n-order chain in this progression.

3.3 Case III: Convergence of the Conversation Chain

In this case we consider our broadest view of communication and examine the dependence structure between the various topics of a conversation. Our state-space consists of observed and diagramed conversations between two or more people as they transition from beginning, through various topics, to their ultimate conclusion. Our intent here is to confirm that the convergence of the chain which models the path of observed conversations can be predicted. In our initial work, we diagramed each conversation from its beginning to its own conclusion or to a point that we considered the conversation to have concluded. In other words, if the conversation did not end on its own, we considered a conversation concluded if it did not transition out of the current topic for several minutes. This standard was purely subjective on the part of the observer. In initially applying this standard, we constructed a reducible Markov chain that contained two communication classes and an absorbing state. As one might guess, the absorbing state was the conclusion of the conversation which we called the "goodbye" state. In our second attempt, we collected new data which excluded the "goodbye" state in order to determine whether another state in the new chain would fill the void left by the exclusion of the "goodbye" state. Our intent here was to evaluate the potential of predicting whether or not a conversation would converge to a specific topic or category of topics. As stated earlier, the conversation chain was constructed by reorganizing more than a dozen topics into seven categories. The structure of our revised chain is presented below, followed by its associated matrix.

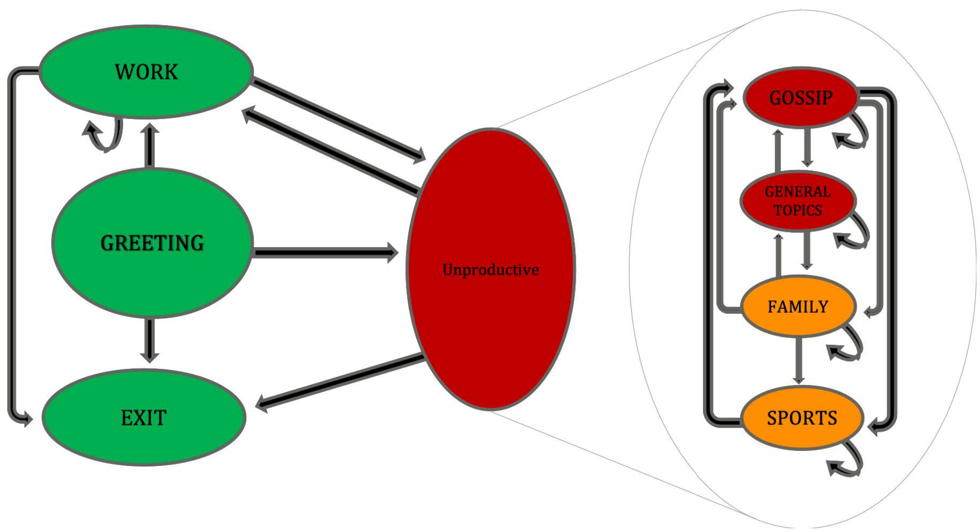


Figure 3.11: Initial conversation chain matrix

P(x,y)	GREETING	WORK	GOSSIP	GENERAL TOPICS	FAMILY	SPORTS	GOODBYE
GREETING	0	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{4}{9}$	0	$\frac{1}{9}$
WORK	0	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	0	$\frac{1}{10}$
GOSSIP	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
GENERAL TOPICS	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$
FAMILY	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$
SPORTS	0	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{4}$
GOODBYE	0	0	0	0	0	0	1

Figure 3.12: Initial conversation chain matrix (labeled)

Our procedure for populating the Conversation Chain matrix began by using the flow chart pictured here to diagram the various conversations that we observed. As this example shows, the conversation started with a *greet*(ing) that lasted approximately 5 seconds. The topic of the conversation then transitioned to *books* and *reading* that lasted for approximately 1 minute and 14 seconds before transitioning to *basketball* and *all sports*. After 1 minute and 45 seconds, the conversation transitioned to

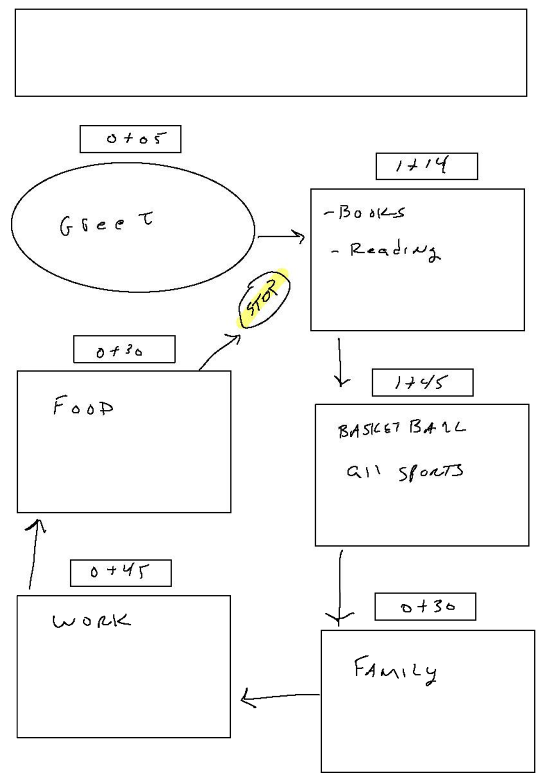


Figure 3.13: Conversation Data Flow Chart

family, work and finally *food* before coming to an end. The duration of the final three topics can be seen in the small rectangular boxes above each listed topic. After collecting a sufficient amount of conversation data, we consolidated it into the seven general topics. To populate our matrix, we tallied the number of times a conversation entered one of our seven general topics. We then tallied to which of the seven topics the conversation transitioned for each of those times. For example, and without loss of generality consider the first topic (*Greeting*) in the matrix below.

P(x,y)	GREETING	WORK	GOSSIP	GENERAL TOPICS	FAMILY	SPORTS	GOODBYE
GREETING	0	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{4}{9}$	0	$\frac{1}{9}$
WORK	0	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	0	$\frac{1}{10}$
GOSSIP	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
GENERAL TOPICS	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$
FAMILY	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$
SPORTS	0	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{4}$
GOODBYE	0	0	0	0	0	0	1

We can see from the denominator in the *Greeting* row that our diagramed conversations included a greeting as a topic on nine different occasions. If we follow the *Greeting* row to the right, we see that in no conversation did the topic transition from a greeting to a greeting. Continuing across the row, by observing the numerators, we can see that of the nine times our conversations included a greeting that they then transitioned twice to *Work*, once to *Gossip*, once to *General Topics*, four times to *Family*, zero times to *Sports*, and once to *Goodbye*. The same relationship of the rows and columns throughout the rest of the matrix creates a Markov Chain on the space defined by the seven conversation topics.

$$P(x,y) = \begin{pmatrix} 0 & \frac{2}{9} & \frac{1}{9} & \frac{1}{9} & \frac{4}{9} & 0 & \frac{1}{9} \\ 0 & \frac{4}{10} & \frac{2}{10} & \frac{2}{10} & \frac{1}{10} & 0 & \frac{1}{10} \\ 0 & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} \\ 0 & \frac{1}{6} & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ 0 & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{2}{8} & \frac{2}{8} \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \implies P^{41}(x,y) \approx \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Convergence of the initial conversation chain matrix

(rounded to four significant figures)

The initial conversation chain converges at 41 iterations from any point in the chain. What differentiates the initial conversation chain from the previous cases is that it is reducible with two communication classes. The first six states have a clear path to communicate with one another while the seventh and absorbing state has no path to any of the first six states. As defined earlier, once the chain enters the absorbing state the chain remains in that state. In layman's terms, once the conversation has ended, the chain has stopped moving (as one would naturally expect). What is noteworthy is that the dependence structure converges without regard to the context of the state space-which distinguishes this case from case II.

In our second conversation chain, we utilized the same procedures defined above however we removed the "goodbye" state to produce the following six-state matrix.

$$P(x,y) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{6}{7} & 0 & \frac{1}{7} \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{4}{17} & \frac{1}{17} & \frac{8}{17} & \frac{1}{17} & \frac{3}{17} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{2}{5} & \frac{1}{5} & 0 \end{pmatrix} \implies \mathrm{P}^{18}(x,y) \approx \begin{pmatrix} 0 & 0.248 & 0.029 & 0.494 & 0.107 & 0.122 \\ 0 & 0.248 & 0.029 & 0.494 & 0.107 & 0.122 \\ 0 & 0.248 & 0.029 & 0.494 & 0.107 & 0.122 \\ 0 & 0.248 & 0.029 & 0.494 & 0.107 & 0.122 \\ 0 & 0.248 & 0.029 & 0.494 & 0.107 & 0.122 \\ 0 & 0.248 & 0.029 & 0.494 & 0.107 & 0.122 \end{pmatrix}$$

Without the absorbing state, we have constructed an irreducible, single communication class chain that quickly converges (rounded to 3 significant figures) to its stationary distribution. Though we failed to indentify another state that would fill the void left by the exclusion of the "goodbye" state. We did confirm that the dependence structure of our second conversation chain does converge to predictable values. With this convergence we can say that, left uninterrupted, the conversation in *this* environment will trend towards the topics and probability associated with each column in the matrix.

CHAPTER 4

Conclusions

4.1 Findings

The history of the Law of Large Numbers provides us with an understanding of what it means to be "Markovian". In disproving Nekrasov's claim that independence is a necessary condition for the Law of Large Numbers, Markov showed that the most essential element of chain dependence is convergence. Without convergence, a given chain's behavior cannot be predicted and therefore cannot be Markovian. Armed with this fact, we presented three cases of varying scope to determine whether or not language is Markovian.

In Case I (Convergence of the VC Chain) we showed that the underlying structure of the VC chain quickly converges to its stationary distribution regardless of the chain's initial distribution. Because the VC chain represents the most fundamental aspect of written language, the context of each state in the chain is not relevant to the chain's utility. Though the VC chain obeys various Markov properties, its only relevant purpose is to verify (using modern methods) Markov's counter-example to Nekrasov's claim.

In Case II (Convergence of the Letter Chain) we considered the 26 individual characters of our corpus. We first evaluated its underlying dependence structure (without regard to context) and showed convergence at the 11^{th} iteration of the chain. But when we considered the context of each state in the chain, we were not able to conclude that the chain converged to predictable values. Therefore, though the underlying dependence structure obeys various Markov properties, the Letter chain has limited utility at this point in our research.

In Case III (Convergence of the Conversation Chain) we evaluated two chains that were similar except for the presence of an absorbing state in the first chain. We once again showed that the underlying dependence structure of each chain converged to predictable values within a reasonable number of iterations. The Conversation chain differs from the Letter chain with regards to context because the Conversation chain has utility to its convergence. Specifically, using the Conversation chain model, one could map and therefore predict the conversation culture of a given environment.

Given the results of Cases I, II, III we conclude that language has an underlying dependence structure that can be modeled and shown to converge to predictable values. This confirms that language can be Markovian. When we consider the context however, we can conclude that the utility of the 'Markovian-ness" of language can be limited by that context.

4.2 Further Research

In chapter one we provided a brief history of the Law of Large Numbers and the inception of Markov chains. Though we were guided by the question of the "Markoivan-ness" of language, we found that the fued between Nekrasov and Markov compelled us to ask a second question. What motivated Markov to expend so much energy disproving Nekrasov's assertion? We developed a conjecture that Markov did so in defense of his friend, teacher and colleague P. L. Chebyshev. We were able to review several primary sources that alluded to Markov's animosity towards Nekrasov. Accordingly, we would like to further examine the relationship between these two men to prove or disprove our conjecture. Specifically, was Markov's relationship to Chebyshev a key motivating factor in his efforts?

In chapter three, we examined the dependence structure of the Letter chain. We found that though the underlying machinery of the chain did behave in a Markovian manner the context of the qualitative random variable has little utility. We showed that by changing the order of the dependence structure we could reproduce entire sentences from our corpus in a random order. We would like to continue evaluating this chain in a manner that would allow us to reproduce our entire corpus in its original form. Our current strategy is to expand the $N-$ order Markov property to a value that requires the chain to consider, by necessity, entire words as "current states" when determining the conditional probability of advancing through the various states of the chain.

In chapter three, we also presented our Conversation chain. We showed that the chain behaves in a Markovian manner in the underlying machinery as well as in the context of the chain. We believe that the concept of our Conversation chain can be extended to accurately predict the "conversation culture" of a given environment. Accordingly, we would like make improvements to the data collection process so that the chain is used to evaluate larger sample sizes. We believe that doing so will enhance the predictive nature of our chain.

Another refinement to our Conversation chain could be the effect of time on its convergence. We showed that we could predict the outcome of a conversation by allowing it to advance through its state-space without interruption. We would like to examine how the length of time spent in each state impacts the chain's progression and ultimate convergence.

LIST OF FIGURES

Bibliography

[1] G. P. Basharin, *The life and work of a. a. markov*, Science Direct (Linear Algebra and its Applications), (2003), pp. 3–26. Linear Algebra and its Applications 386(2004) 3-26.

[2] P. Butzer and F. Jongmans, *P. l. chebyshev (1821-1894) a guide to his life and work*, Journal of Approximation Theory, 96 (1999), pp. 111–138. Article ID jath.1998.3289.

[3] S. R. Dunbar, *Topics in probability theory and stochastic processes.* downloaded from https://www.math.unl.edu/ sdunbar1/ProbabilityTheorey/Lessons/BernoulliTrials/DemoivreLaplaceCLT /demoivrelaplaceclt.pdf on 2/2/19.

[4] I. Grzegorczyk, *Mathematics and Fine Arts*, Kendall/Hunt Publishing Co., 2000.

[5] B. Hayes, *First links in the markov chain*, American Scientist, 101 (2013). March-April.

[6] P. HOEL, *Introduction to Stochastic Processes*, Houghton Mifflin Company, Boston MA, 1972.

[7] D. LINK, *Traces of the mouth: Andrei andreyevich markov's mathematization of writing*, Science History Publications Ltd, (2006). 0073-2753/06/4403-0321.

[8] A. A. MARKOV, *An example of statistical investigation of the text eugene onegin concerning the connection of samples in chains*, Science in Context, (2006), pp. 591–600. DOI:10.1-17/S0269889706001074.

[9] K. O. ONDAR, *The Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*, Springer-Verlag, New York, New York, 1981.

[10] E. SENETA, *Markov and the birth of chain dependence theory*, International Statistical Review, 3 (1996), pp. 255–263. International Statistical Review(3) 1966,64,255-263.

[11] E. SENETA, *A tricentenary history of the law of large numbers*, Benoulli, 19 (2013), pp. 1088–1121. DOI: 10.3150/12-BEJSP12.

[12] S. M. STIGLER, *The History of Statistics: The Measurement of Uncertainty before 1900*, The Belknap Press of Harvard University Press, Cambridge, MA and London, England, 1986.

Non-Exclusive Distribution License

In order for California State University Channel Islands (CSUCI) to reproduce, translate and distribute your submission worldwide through the CSUCI Institutional Repository, your agreement to the following terms is necessary. The author(s) retain any copyright currently on the item as well as the ability to submit the item to publishers or other repositories.

By signing and submitting this license, you (the author(s) or copyright owner) grants to CSUCI the nonexclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that CSUCI may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that CSUCI may keep more than one copy of this submission for purposes of security, backup and preservation.

You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. You also represent and warrant that the submission contains no libelous or other unlawful matter and makes no improper invasion of the privacy of any other person.

If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant CSUCI the rights required by this license, and that such third party owned material is clearly identified and acknowledged within the text or content of the submission. You take full responsibility to obtain permission to use any material that is not your own. This permission must be granted to you before you sign this form.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN CSUCI, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

The CSUCI Institutional Repository will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

The Convergence of Markov Chains and Language

Title of Item

Markov Chains, Language, Law of Large Numbers

3 to 5 keywords or phrases to describe the item

Thomas L. LaGrange

Author(s) Name (Print)

12/9/19

Author(s) Signature Date