Markov Chains and Its Applications to Golf

by
Wade Shasky

A project submitted to the Department of
Mathematical Sciences in conformity with the requirements
for Math 4301 (Honour's Seminar)

Lakehead University
Thunder Bay, Ontario, Canada

Abstract

The purpose of this paper is to look at and research Discrete-Time Markov Chains. An in-depth description of various properties a Markov Chain can have will be viewed. This includes the basics of discrete-time Markov Chains and more challenging concepts that further describe a Markov Chain. Then, the research and knowledge gained on Markov Chains will be applied to the game of golf. The application will try to correctly pick the winner of a match-play, and stroke-play, golf event using the theory of Markov Chains. A final application using correlation will show the best way golfers can improve their scores.

Acknowledgements

This honours project would not have been possible without the help and support of many people. First of all, I would like to thank the supervisor of the Honours Seminar course, Professor Razvan Anisca, for the expertise and guidance he has given throughout the year. I would also like to thank the PGA Tour for providing the information and statistics used in the application section of this paper. I would like to express my greatest appreciation to Professor Wendy Huang. The help and direction she provided me for this project is nothing short of amazing. Without the knowledge and hours of help from Professor Huang, this paper would not be what it is. Lastly, I would like to thank my parents. Without their love and support they have provided me throughout my life, none of this would ever have been possible.

Contents

CHAPTER 1

Introduction

The goal of this project is to investigate a mathematical property, called Markov Chains and to apply this knowledge to the game of golf. In order to understand the theory of Markov Chains, one must take knowledge gained in Linear Algebra and Statistics.

Markov Chains, named after the Russian mathematician Andrey Markov, is a type of stochastic process dealing with random processes. There are two types of Markov Chains; Discrete-Time, a countable or finite process, and Continuous-Time, an uncountable process. The scope of this paper deals strictly with discrete-time Markov Chains. These are a stochastic process that satisfies three properties:

(1) Discrete-Time
(2) Countable or finite state space
(3) Future location depends only on the present

These properties and definitions will be clearly presented along with other properties a Markov Chain can have.

Once discrete-time Markov Chain theory is presented, this paper will switch to an application in the sport of golf. The most elite players in the world play on the PGA Tour. This paper will use the knowledge and theory of Markov Chains to try and predict a winner of a match-play style golf event. This is where two golfers go head-to-head and compete on a hole-to-hole basis and whomever has won more holes at the end of 18 holes is declared the winner. A winner in a stroke-play event will also be predicted. Then, I will examine the best way for a player on the PGA Tour to improve their score by doing a correlation model of multiple statistics.

This paper is structured as follows. In Chapter 2, Markov Chains will be explained. This will start with basic definitions and properties a Markov Chain must have. Then, matrices and digraphs will be introduced in order to best express a Markov Chain. Finally, more in-depth properties to better classify Markov Chains will be described. Chapter 3 will take the knowledge described in Chapter 2 and apply it to golf. First, the game of golf will be explained with terms and examples. Then a winner will be predicted in a match-play style of golf event, then a stroke-play event, when two players from the PGA Tour will face off against each other. Chapter 4 will study various statistics PGA Tour players have and a correlation model will be used to determine the best way a player can improve their game in order to lower their score.

CHAPTER 2

Discrete-Time Markov Chains

1. Introduction

In this chapter, the fundamental concepts of Markov Chains, to be more specific, discrete-time Markov Chains, will be reviewed. This will create a foundation in order to better understand further discussions of Markov Chains along with its properties and applications. Most of the information here was taken from [**1**] with help on examples from [**4**]. To begin our discussion of Markov Chains, stochastic processes must first be introduced. Before that though, some basic terms will be defined.

Definition 2.1.1: A *random variable* is a function that maps the set of all possible outcomes in an experiment into the real numbers, $\mathbb{R}$.

Definition 2.1.2: Random variables are considered *countable* when the number of variables being considered can be counted. The converse is when the number of random variables is unknown or never ends, then it is considered *uncountable.*

Definition 2.1.3: A *finite* set is one that has limits on the amount of variables or outcomes possible.

We are now ready to define a stochastic process and begin our discussion of Markov Chains.

Definition 2.1.4: A *stochastic process* is a collection of random variables transfixed in time defined by a set of possible outcomes. If the collection of random variables is countable or finite,$\{X_k : k = 0, 1, 2, ...\}$, then it is called a *discrete-time process.* If the collection of random variables is uncountable,$\{X_t : t \geq 0\}$, then it is called a *continuous-time process.*

Definition 2.1.5: The set of distinct values, those different from one another, assumed by a stochastic process is called *state space.* If the state space of a stochastic process is countable, or finite, then the process is called a *chain.*

Based on these definitions, this paper will only consider discrete-time stochastic processes, meaning that the process of the state space will be a chain.

The distinction between stochastic processes and Markov Chains is further clarified by the Markov Property.

Definition 2.1.6: A stochastic process $\{X_k : k = 0, 1, 2, ...\}$ with a state space of $S = \{1, 2, ...\}$ is said to satisfy the *Markov Property* if for every k and all states $\{i_1, i_2, ..., i_k\} \in S$ it is true that

$$P[X_k = i_k \mid X_{k-1} = i_{k-1}, X_{k-2} = i_{k-2}, ..., X_1 = i_1] = P[X_k = i_k \mid X_{k-1} = i_{k-1}]$$

This means that a particle, which is defined by any point or number in the state space, is in state i at time n. In other words, the Markov Property is satisfied if the future location of the particle depends on its present location, NOT its past. With all this information we are ready to define a discrete-time Markov Chain.

Definition 2.1.7: A *discrete-time Markov Chain* is a stochastic process that must satisfy the following restrictions:

(1) Discrete-Time
(2) Countable or finite state space
(3) The future location depends on the present, not its past (Markov Property)

We do however, need to know if time depends on when a transition from state i to state j actually takes place.

Definition 2.1.8: A discrete-time Markov Chain is *stationary* in time if the probability of going from one state to another is independent of the time in which the step is being made. We can express this for states i to j as

$$P[X_n = j \mid X_{n-1} = i] = P[X_{n+k} = j \mid X_{n+k-1} = i] \ \ for\ k = -(n-1), -(n-2), ..., 0, 1, 2, ...$$

If this property fails, we consider the discrete-time Markov Chain *non-stationary.*

For this paper, we will only consider and study stationary discrete-time Markov Chains.

2. Transition Matrices

Now we need to describe a Markov Chain as a probability since we are interested in the odds of moving from one state to another. Since we are interested in a state i and j at time n, we can define the state at time n as the conditional probability given by $P[X_n = j \mid X_{n-1} = i]$ for all n because it is stationary. This is the probability of going from state i to state j in n steps which is defined next.

Definition 2.2.1: A *transition probability* is the probability of moving from one state, say i, to another, say j, in a discrete number of n steps, denoted as $p_{ij}^{(n)}$.

With this information we need a way to represent discrete-time stationary Markov Chains with finite state space, $\{X_k\}$. In order to represent state i and state j together, the most convenient form of doing so will be used, which is a matrix, P. By associating the i^{th} row and column of P with the i^{th} state of S and similarly with state j, a transition probability matrix of the form p_{ij} will be achieved as follows

$$P = \begin{bmatrix} p_{11} & p_{12} & .. & p_{1n} \\ p_{21} & p_{22} & .. & p_{2n} \\ : & : & .. & : \\ p_{n1} & p_{n2} & .. & p_{nn} \end{bmatrix}$$

Proposition 2.2.2: Every transition matrix has the following properties:

(1) all entries are non-negative
(2) the sum of the entries in each row is one

Proof:

(1) $p_{ij} \geq 0$ since probabilities can only be represented as positive values
(2) For any $i = 1, 2, ..., n$

$$\begin{aligned} & p_{i1} + p_{i2} + ... + p_{in} \\ & = P[X_k = 1 \mid X_{k-1} = i] + P[X_k = 2 \mid X_{k-1} = i] + ... + P[X_k = n \mid X_{k-1} = i] \\ & = P[(X_k = 1) \cup (X_k = 2) \cup ... \cup (X_k = n) \mid X_{k-1} = i] \\ & = P[X_k \in S \mid X_{k-1} = i] \\ & = 1 \end{aligned}$$

□

Defintion 2.2.3: A *transition matrix* is a matrix that satisfies Proposition 2.2.2.

Another way of representing Markov Chains is with a digraph. A digraph is commonly used in graph theory and just gives a more visual way of expressing a Markov Chain and can make classification of Markov Chains easier. Essentially, digraphs take the information in a matrix and then maps the rows, or state i, to the columns, state j, based on the values from each state. The following example will illustrate this and transition matrices.

Example 2.2.4: Let x represent the voting results of an election with the three major parties represented as a vector with State Space $S = \{C, L, N\}$ as:

$$x = \begin{bmatrix} \%\text{ Voting Conservative} & \text{(C)} \\ \%\text{ Voting Liberal} & \text{(L)} \\ \%\text{ Voting N.D.P.} & \text{(N)} \end{bmatrix}$$

Now suppose we record the election results in back-to-back elections as a stochastic process to describe the voting changes from one election to the next. An example of a transition matrix representing this data is:

$$\begin{bmatrix} & & \text{TO} & & \\ & & \text{C} & \text{L} & \text{N} \\ & \text{C} & 0.80 & 0.15 & 0.05 \\ \text{FROM} & \text{L} & 0.25 & 0.50 & 0.25 \\ & \text{N} & 0.10 & 0.20 & 0.70 \end{bmatrix}$$

The entries in the first row, labeled C, describe what the people who voted Conservative in one election did the next. In this particular set of hypothetical data, 80% will vote Conservative again, 15% will change their vote to Liberal, and 5% will change their vote to N.D.P. The other rows, L and N, have similar interpretations.

Note that this is a 3×3 transition matrix since there are no negative values and the sum of the entries in each row is 1.

For this transition matrix, the corresponding digraph is shown below:

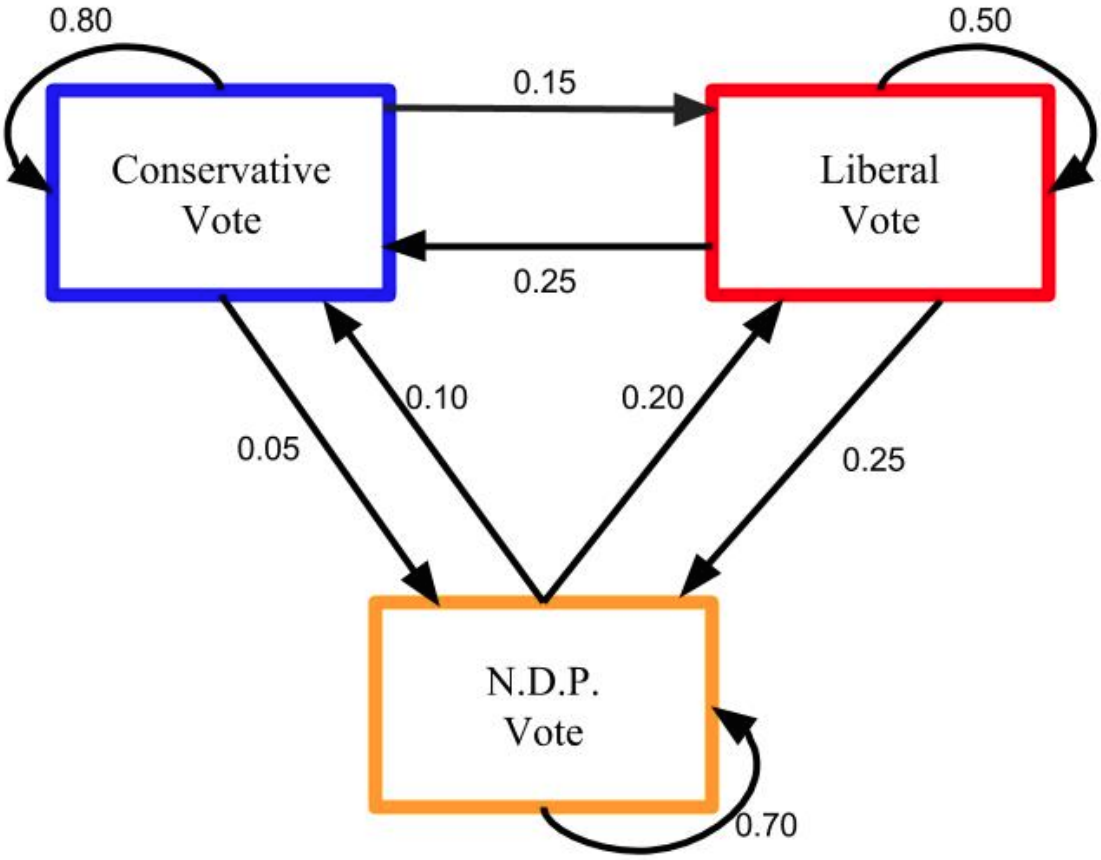

Figure 1

Definition 2.2.5: Any square $n \times n$ matrix that satisfies Proposition 2.2.2 is called a *stochastic matrix.*

Here an important note will be made. A transition matrix is usually referred to only when a specific Markov Chain is being considered. However, when we refer to a stochastic matrix, there is no attempt to relate the data to a specific Markov Chain. The reason this is important is because when developing the theory of Markov Chains, we simply consider stochastic matrices and apply the theory to a Markov Chain and use a transition

matrix. This distinction is important because the theory behind Markov Chains deals with stochastic matrices. In contradiction, when we are considering a Markov Chain we just apply the theory of stochastic matrices to obtain a transition matrix. Once it is known a discrete-time stationary Markov Chain is what we want to consider, then we find the corresponding transition matrix.

Next, we want to determine where a Markov Chain is at any time and in order to do this, we must first know where the chain started.

Definition 2.2.6: A vector $\mathbf{a}_0 = (\alpha_1, \alpha_2, ...\alpha_n)$ is called a *starting vector* if we have $\sum_{i=1}^{n} \alpha_i = 1$ and $\alpha_i \geq 0$ for $i = 1, 2, ...n$.

When a Markov Chain process starts at $\mathbf{a}_0$, which is a single state with certainty, it has a one in the coordinate corresponding to that state and zeros elsewhere. For convenience, the starting vector refers to the distribution at time zero, denoted by $\alpha_k = P[X_0 = k]$, for $k = 1, 2, ..., n$.

By using conditional probability, we can find

$$\begin{aligned} P[X_1 = i] &= P[X_0 = 1] P[X_1 = i \mid X_0 = 1] + P[X_0 = 2] P[X_1 = i \mid X_0 = 2] \\ &\quad +... + P[X_0 = n] P[X_1 = i \mid X_0 = n] \\ &= \sum_{j=1}^{n} \alpha_j P_{ji} \end{aligned}$$

Now we can express this in terms of our transition matrix P as an expression $P[X_1 = i]$ of the i^{th} coordinate of the vector $\mathbf{a}_0 P$, denoted by $\mathbf{a}_1$. Furthermore, $\mathbf{a}_2 = (\mathbf{a}_0 P) P = \mathbf{a}_0 P^2$ where the i^{th} coordinate of $\mathbf{a}_2$ is $P[X_2 = i]$.

Proposition 2.2.7: The distribution of where a process is after n steps with starting vector $\mathbf{a}_0$ is given by $\mathbf{a}_n = \mathbf{a}_0 P^n$, where P^n is the transition matrix represented by the n step transition probabilities.

This means that we can write the elements of P^n as $p_{ij}^{(n)} = P[X_{k+n} = j \mid X_k = i]$, for any $n > 1$.

Example 2.2.8: Once again consider the transition matrix P given in Example 2.2.4

$$P = \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.25 & 0.50 & 0.25 \\ 0.10 & 0.20 & 0.70 \end{bmatrix}$$

If the data is repeated for another election, $n = 2$, simple matrix multiplication will yield the following election results, interpreted in a similar manner as before:

$$P^2 = P \times P = \begin{bmatrix} 0.682\,5 & 0.205 & 0.112\,5 \\ 0.35 & 0.337\,5 & 0.312\,5 \\ 0.2 & 0.255 & 0.545 \end{bmatrix}$$

Note that since the calculations of P^n can be very difficult for larger matrices and can become tedious to compute, for larger calculations, a computer will be used.

This brings us to a very important theorem in Markov Chain theory.

Theorem 2.2.9 (Chapman-Kolmogorov Identity): For all non-negative integers l and m

$$p_{ij}^{(l+m)} = \sum_{k \in S} p_{ik}^{(l)} p_{kj}^{(m)}$$

Proof:

$$\begin{aligned} p_{ij}^{(l+m)} &= P\left[X_{l+m} = j \mid X_0 = i\right] \\ &= \sum_{k \in S} P\left[X_{l+m} = j,\ X_l = k \mid X_0 = i\right] \text{ By countable additivity} \\ &= \sum_{k \in S} P\left[X_{l+m} = j \mid X_l = k,\ X_0 = i\right] P\left[X_l = k \mid X_0 = i\right] \\ &= \sum_{k \in S} P\left[X_{l+m} = j \mid X_l = k\right] P\left[X_l = k \mid X_0 = i\right] \text{ By the Markov Property} \\ &= \sum_{k \in S} p_{ik}^{(l)} p_{kj}^{(m)} \qquad \square \end{aligned}$$

This identity is important since it tells us the probability of getting from state i to j in $l + m$ steps is the same as if we go from i to k in l steps then to j in m steps. In other words it does not matter what path we take to get from one state to another, as it has the same probability. This identity is very important in the application section.

One application of the Chapman-Kolmogorov Identity is that the (i, j) entry of P^n is equal to $p_{ij}^{(n)}$. This means that the matrix P^n is a transition matrix that represents the n step transition probabilities, which helps in the understanding of Proposition 2.2.7. as it is possible because of the Chapman-Kolmogorov Identity.

Definition 2.2.10 [4]: A scalar λ is called an *eigenvalue* of a matrix P if there is a nontrivial vector $\mathbf{x}$ of $P\mathbf{x} = \lambda\mathbf{x}$. Such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.

A scalar λ is an eigenvalue of an $n \times n$ transition matrix P if and only if λ satisfies the characteristic equation $\det(P - \lambda I) = 0$, where I is the identity matrix. Thus, in order to find eigenvalues, the characteristic equation must be used.

I will end this section with an example of how to calculate eigenvalues [**4**]. These are used to describe a specific value that a matrix can achieve and is important in understanding matrix theory. All the information in this section is important as it laid the ground work for understanding the basics of Markov Chains, which is critical in understanding and being able to accomplish the applications of Markov Chains to the sport of golf.

Example 2.2.11: Use the characteristic equation to find the eigenvalues of the matrix $P = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$. Based on the previous definition and the characteristic equation we find that

$$\det(P - \lambda I) = \det\left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \det\begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = 0$$

So by definition of the determinant we have that

$$\begin{aligned} \det(P - \lambda I) &= (2-\lambda)(-6-\lambda) - (3)(3) \\ &= -12 + 6\lambda - 2\lambda + \lambda^2 - 9 \\ &= \lambda^2 + 4\lambda - 21 \\ &= (\lambda - 3)(\lambda + 7) \end{aligned}$$

Since $\det(P - \lambda I) = 0$, then $\lambda = 3$ or $\lambda = -7$. Thus the eigenvalues of P must be 3 and -7.

3. Properties and Classification

In this section some more properties a Markov Chain can have will be discussed [**1**]. This will allow us to classify Markov Chains even further. I will begin this section with a definition but it is important to keep digraphs in mind as they will help in the understanding of some of the classifications a Markov Chain can have.

Definition 2.3.1: Let S be a Markov Chain. A subset, C, of state space, S, is called *closed* if $p_{ij} = 0$ for all $i \in C$ and $j \notin C$. If a closed set consists of a single state, then that state is called an *absorbing state*.

An important note must be made here about a subset being closed. In Markov Chains, being closed does not refer to the notion from topology. Here when the subset is closed it is referring to the impossibility of leaving. Thus, a subset of S is closed if once the chain enters C, it can never leave.

Example 2.3.2: Recall Example 2.2.3. The transition matrix P is not closed or absorbing, since there is always a way to change your vote from one party to another in the next election, and all the votes do not converge to one election party respectively.

Example 2.3.3: Consider the following digraph:

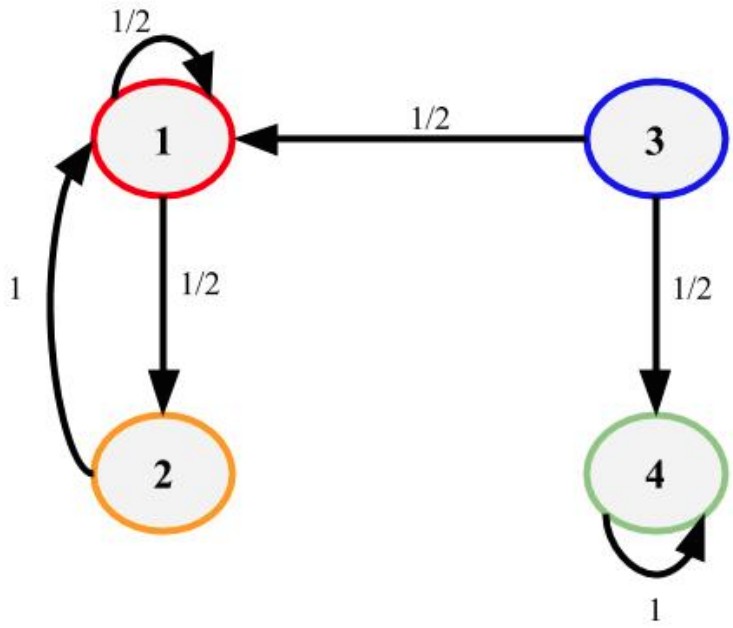

Figure 2

It is clear that state 4 is absorbing since once the chain enters, it can never leave. Also note that $\{1, 2\}$ is closed since once this state is entered, nothing can escape it.

Definition 2.3.4: A Markov Chain is called *irreducible* if there exists no non-empty closed set other than S itself. If S has a proper closed subset, it is called *reducible*.

Definition 2.3.5: If two states, i and j satisfy, for some $n > 0$, $p_{ij}^{(n)} > 0$ and for some $m > 0$, $p_{ji}^{(m)} > 0$ then the states i and j are said to *intercommunicate*.

This definition tells us it is possible to go from state i to j in n steps and from state j to i in m steps even if $n \neq m$, which is very important when applying Markov Chains to golf. Next, we will relate the last two definitions together to achieve the following theorem.

Theorem 2.3.6: A Markov Chain is irreducible if and only if all pairs of states intercommunicate.

Proof:

Assume the chain is irreducible. Define $C_j = \left\{ i : p_{ij}^{(n)} = 0 \text{ for all } n \geq 0 \mid i \neq j \right\}$. That is, C_j is the set of all states from which state j cannot be reached. Assuming $C_j \neq \emptyset$ we will show that C_j is closed by showing that if $i \in C_j$ and $k \notin C_j$, then $p_{ik} = 0$.

Now if $k \notin C_j$ then for some $m \geq 0$ it follows that $p_{kj}^{(m)} > 0$. If p_{ik} were positive, then $p_{ij}^{(m+1)} = \sum_{l \in S} p_{il} p_{lj}^{(m)} \geq p_{ik} p_{kj}^{(m)} > 0$, which implies $i \notin C_j$. This contradiction leads us to conclude that $p_{ik} = 0$ for all $i \in C_j$, $k \notin C_j$, so C_j is closed. The only nonempty closed subset of an irreducible chain is S so $C_j = S$. However $j \notin C_j$ since $p_{jj}^{(0)} = 1$, giving a contradiction. Therefore $C_j = \emptyset$ which means j can be reached from all states. Since j is an arbitrary state we have that all states intercommunicate.

Conversely, assume that all states intercommunicate. Let C be a nonempty closed set in S. If $j \in C$, then for an arbitrary state $i \in S$, there exists an n_i such that $p_{ji}^{(n_i)} > 0$. Since we have that state i can be reached from state $j \in C$, it follows that $i \in C$. However i was an arbitrary state in S, thus $C = S$. Therefore, it must be the case that the chain is in fact irreducible. □

Example 2.3.7: Consider a Markov Chain with the following transition matrix and corresponding digraph:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.6 & 0 & 0.4 \\ 1 & 0 & 0 \end{bmatrix}$$

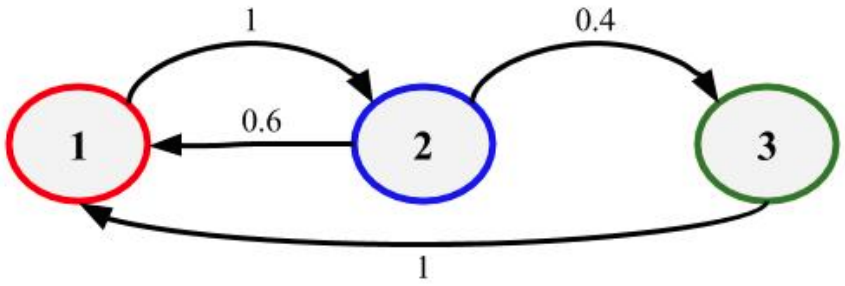

Figure 3

If $S = \{1, 2, 3\}$ then states 1 and 2 intercommunicate since $p_{12}^{(1)} > 0$ since the probability of going from state 1 to 2 occurring in 1 step is 1. Also, $p_{21}^{(1)} > 0$ and $p_{21}^{(2)} > 0$ since the probability of going from state 2 to 1 in 1 step is $0.6 > 0$ and the probability of going from state 2 to 1 in 2 steps is $(1)(0.4) = 0.4 > 0$. Further note that all states in S intercommunicate, thus by theorem 2.3.6, this chain is irreducible.

Definition 2.3.8: A state i has *period* d if the following two conditions hold
(1) $p_{jj}^{(n)} = 0$ unless $n = md$ for some positive integer m and
(2) d is the largest integer with property 1.

If $d = 1$ then the state is said to be *aperiodic.*

The easiest way to calculate the period of a state is by following the next theorem.

Theorem 2.3.9: State j has period d if and only if d is the greatest common divisor of all of those n's for which $p_{jj}^{(n)} > 0$. That is

$$d = G.C.D. \left\{ n : p_{jj}^{(n)} > 0 \right\}$$

The period of state j is concerned with the times at which the chain might possibly return to state j. The next part of this paper will explore the properties of Markov Chains and when and if they can return to some state.

Definition 2.3.10: The first visit to state j from state i occurs at time n is defined as a probability, $f_{ij}^{(n)}$. In other words,

$$f_{ij}^{(n)} = P\left[X_{n+k} = j \mid X_{n+j-1} \neq j, X_{n+k-2} \neq j, ..., X_{k+1} \neq j, X_k = i\right].$$

We refer to $f_{ii}^{(n)}$ as the probability that the first return to state i occurs at time n when $i = j$. By definition, we say $f_{ij}^{(0)} = f_{ii}^{(0)} = 0$.

Definition 2.3.11: Let $f_{ij}^{*} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ for fixed states i and j, where th probability of ever visiting state j from i is f_{ij}^{*}. If $i = j$ then $f_{ii}^{*} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$, which denotes the probability of eventually returning to state i.

Definition 2.3.12: A state j is *persistant* if $f_{jj}^{*} = 1$. If $f_{jj}^{*} < 1$, then j is considered to be *transient* and we define the expected return time to state j as

$$\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}.$$

The next example will help to make these concepts a little more clear.

Example 2.3.15:

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

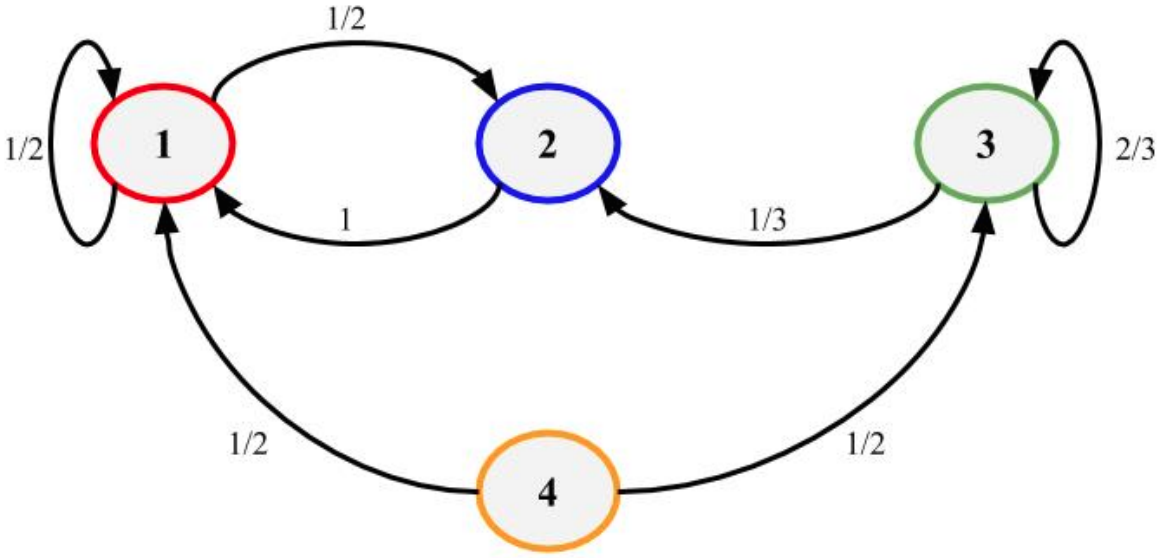

Figure 4

$$
\begin{aligned}
f_{44}^{(n)} &= 0, \text{ for all } n \text{ so state 4 is transient} \\
f_{33}^{(1)} &= 2/3, \ f_{33}^{(n)} = 0 \text{ for } n \geq 2 \text{ so state 3 is transient} \\
f_{11}^{*} &= f_{11}^{(1)} + f_{11}^{(2)} = 1/2 + 1/2 = 1 \\
f_{22}^{*} &= f_{22}^{(1)} + f_{22}^{(2)} + ... + f_{22}^{(k)} + ... = 0 + 1/2 + 1/4 + 1/8 + ... = 1 \\
\text{Then } \mu_1 &= \sum_{k=1}^{\infty} k f_{11}^{(k)} = 1 \cdot 1/2 + 2 \cdot 1/2 = 3/2 \\
\mu_2 &= \sum_{k=2}^{\infty} k f_{22}^{(k)} = \sum_{k=2}^{\infty} \frac{k}{2^{k-1}} = 3
\end{aligned}
$$

Thus both states 1 and 2 are positive persistant.

This ends our discussion on the theory of Markov Chains. The classification and properties that a Markov Chain can have will now be used in an application to the sport of golf.

CHAPTER 3

Application to Golf

1. Introduction to Golf

In this chapter, the sport of golf will be explained through terms, definitions and explanations needed to understand the basics of golf. This will create a foundation of knowledge about the sport in order to understand the application part of this paper. Once this is done, Markov Chain theory described in the previous chapter will be used to predict a winner of a golfing event.

This section will be an introduction to golf to create a full background of knowledge needed for the next section where a winner of a golfing event is attempted to be predicted.

To start, a brief history of the sport will be discussed. It is said that golf has been played for 100's of years but the golf played today is of Scottish invention. The first game of golf ever played was in 1552 at the Old Course at St. Andrew's in Scotland. The first set of rules of golf were made in 1744. An old wives' tale as to why a golf course is 18 holes, established in 1764 down from the original 22, is because it took 18 holes to finish a fifth of scotch but the true reason is debated. In 1860, the first professional major championship was played in Scotland. In 1894, the United States Golf Association (USGA), which is what acts as the governing body of golf in North America still to this day was founded. In 1901, the Professional Golfers Association (PGA) was formed [**6**].

Golf is a sport played outside on a golf course. Every golf course varies in distance but the standard course consists of 18 holes. The completion of 18 holes is considered a round. The goal of golf is to hit a golf ball into each of the 18 holes on the golf course. Each hole has its own individual starting point, called the tee shot, and finishing point, the area called a green where a player putts their ball into the hole. Each hole also has its own distance, how far it is from the tee shot to the hole, and specific par. The **score** an accomplished player is expected to make on a hole is called the *par* of a hole.

In other words, each hole on a golf course is assigned a par to which a player should achieve. The par of a hole can be a par 3, 4, or 5. Most courses range from a total par of about 68-73 depending on the number of par 3's, 4's, and 5's but most courses have an overall par of 72.

Definition 3.1.1 [**7**]**:** A *stroke* is the forward movement of the club made with the intention of striking at and moving the ball. The intention of hitting the golf ball happens when a player *addresses the ball* which happens when a player puts their club on the ground either directly in front of or behind the golf ball.

This is how a players score is determined. Thus, one wants to make as little strokes as possible during a round. A players guideline to how many strokes should be taken to get the ball in the hole is the par of that hole. However, this does not always occur; sometimes for better and sometimes for worse.

There are many different scores a player can have on a hole that describe how well they have done in relation to the par of that hole. An *eagle* is a score of two-under-par on a hole. A *birdie* is a score of one-under-par on a hole. A *par* is a score of even on a hole. A *bogey* is a score of one-over-par and a *double bogey* is a score of two-over-par on a hole respectively.

There are more terms and scores a player can attain on a hole but these are the most common scores a player achieves. Since scoring the lowest possible score is best in golf, an eagle is better than a birdie, which is better than a par and so on. Thus if a player scores a birdie on a par-4, they achieved a 3, also known as -1.

The *handicap* of a golfer is used to represent a golfers potential playing ability, with a lower value one achieves representing the better they are. How this numerical value is found for individuals is not important since this paper will apply only to professional golfers, who all have the lowest possible handicap of zero, thus making the playing field even. This happens even though some professional golfers are clearly better than others. Furthermore, this paper will look at golfers who mainly play on the Professional Golfers Association Tour (PGA Tour). The most extensive statistics are kept on this tour, it is the main tour played in North America, and is considered to have the best golfers in the world. This includes the likes of former great Jack Nicklaus, the best known golfer in the world, Tiger Woods, and the worlds number one player at the moment, Rory McIlroy. The PGA Tour also acts under the set of rules and regulations presented by the USGA, stated earlier.

Now I will explain the two main formats in which golf is played.

Definition 3.1.2 [7]**:** *Stroke Play* is a form of competition based on the cumulative number of strokes taken, either over one or several rounds.

Stroke Play is the most used and popular format of golf. Almost all events on the PGA Tour is a stroke play format where the winner of an event is the player who has the lowest cumulative score after four rounds of golf on the same course at the same event. The total score is in relation to the par of the course. This means that if a player shoots a 67 on a par-72 course, they are 5-under par.

The next format in which golf can be played is the main focus of the application section of this paper and we will be interested in this style of play.

Definition 3.1.3 [7]**:** *Match Play* is a competition format in which the round is played with the goal of winning individual holes. Scoring is kept by comparing the number of holes won by each player. Moreover, the final score reflects the margin of victory and at which hole the match ended.

In other words, match-play is a head-to-head competition where each player tries to score the best on each hole in an attempt to win said hole. As an example, we have Player A and Player B. If Player A shoots a par 4 and Player B shoots a birdie 3 on a hole, then Player B wins the hole. At this point, Player B is said to be **'1-up'** and Player A is **'1-down'**. Let's say on the next hole, a par-5, both players shoot a birdie 4. Then this hole is said to be **halved** and the score stays the same. Then at the next hole, assume Player A shoots a par 3 and Player B shoots a double bogey 5. Then Player A has won this hole and the match is now **'all-square'**. This head-to-head competition will continue until a winner is found or at the end of 18 holes. If the players are **'all-square'** after the 18 holes, then the match is **tied**. Let's say that Player A has won 6 holes, lost 5 and tied 7 after 18. Then Player A has won the match by a score of **'1-up'**. In a different scenario, imagine after 15 holes Player A is **'3-up'**. This means that Player B is in a *dormie* situation, meaning that Player B has to win the next three holes to tie the match. If hole 16 is **halved**, then the match is over since Player A is **'3-up'** with only two holes left to play making it impossible for Player B to have a chance; hence, Player B has lost. We say that Player A won by a score of **'3-and2'** meaning Player A was **'3-up'** with two holes to play. By this standard, the shortest possible game could be ten holes.

The match-play format is of interest because it is played on the biggest international competition in the world, and possibly the most exciting stage in golf, The Presidents' Cup and the Ryder Cup.

2. Golf Statistics

Now that an understanding of the basics of golf and formats in which it can be played has been explained, I will turn my attention to determine which PGA Tour players are better than another and by how much using different statistics and rankings on the PGA Tour website [**8**]. The statistics that will be used are based on the 2014 season. The data was tracked using 'ShotLink', which is an online database partnered with the PGA Tour where many statistics have been generated over the year.

Definition 3.2.1 [**7**]**:** The official *World Golf Ranking* is a system for rating the performance level of professional male golfers.

This system is a way to establish where in the world professional golfers are in comparison to others. This is a way of expressing who is considered the best golfer in the world at a specific time and who is the second best and so on. Each player has a specific number of average points gained from each week they compete in a tournament. For every tournament players perform in, they are given a certain world rating value, thus, they can differ from week-to-week. Official events from the leading professional tours from around the world are eligible for ranking points which are awarded according to the tournaments strength of field and players finishing position. How these points are awarded is irrelevant to this paper.

The reason why the World Golf Ranking system is brought up is to give a comparison of how different players are 'officially' ranked in regards to others. This is one way to see who is better than another and is the most widely used and accepted form of determining who is considered the best golfer in the world each week.

Another way of determining who is a better golfer is by using statistics on the PGA Tour website from the 2014 season. There are hundreds of different statistics but in determining who is better and by how much, I will use the following statistics, taken and defined on [**8**], as I feel they are the most indicative as to the performance of an individual player.

(1) *Score Average*: The average number of strokes per completed round taken over a desired time period. In our case, this time period is the 2014 season.
(2) *Birdie Average*: The average number of birdies made per round played.
(3) *Driving Distance*: The average number of yards per measured drive from the tee block. These drives are measured on only two holes per round and measured to the point they come to rest regardless of whether they are in the fairway or not.
(4) *Driving Accuracy*: Expressed as a percentage, it is the number of times a tee shot comes to rest in the fairway (the ideal area a player aims for because it is closely mown making the next shot easier), not including par threes.
(5) *Greens in Regulation(GIR)*: Expressed as a percentage, it is the number of times a player was able to hit a green in regulation, when any portion of the ball is touching the putting surface, divided by the number of holes played. The GIR stroke is determined by subtracting two from par, i.e. a GIR on a par-3 is when the tee shot lands on the green.
(6) *Birdie or Better Conversion*: Expressed as a percentage, it is the number of times a player makes a birdie after hitting a Green in Regulation.
(7) *Scrambling*: Expressed as a percentage, it is the number of times a player misses the Green in Regulation but still is able to make a par or better.

To help give a better idea of these statistics I will give a little example. Let's say that Player A goes out and plays five holes of golf, where the overall par of the holes is 21.

Hole One is a par-4 and the tee shot is hit 300-yards into the fairway. The second shot is then hit and lands on the green, 30-feet from the hole. Here, Player A has made a green in regulation. Player A goes on to miss his first putt but makes the next one from 5-feet away so he scores a par and is **'even'**.

Hole Two is a par-5 and the tee shot is hit 322-yards into the left rough. The second shot is hit to the fairway 102-yards from the hole. The third shot is hit 10-feet from the hole, giving him a GIR, and Player A makes the putt for a birdie 4 and is now **'1-under'**.

Hole Three is a par-4 and Player A hits the tee shot 267-yards into the fairway. The next shot is hit to a bunker beside the green. The bunker shot is hit out to 16-feet from the hole. This putt is missed but the next one is made. Thus Player A shot a bogey 5 and is now back to **'even'**.

Hole Four is a par-3 and the tee shot is hit 176- yards into the rough short of the green. An excellent pitch shot is hit to 2-feet from the hole . The putt is made for a par 3 and Player A is still **'even'**.

Finally, on Hole Five, a par-5, the tee shot is hit 298- yards into the fairway. The next shot is hit to 150-yards from the hole and then the next shot is hit onto the green, for a GIR, 21-feet away. Player A makes this putt for a birdie 4 and has finished **'1-under'**.

Hole	Tee Shot	2nd Shot	3rd Shot	4th Shot	5th Shot
One	300-yards into Fairway	On Green, 30-feet from the hole	Misses putt from 30-feet	Makes Putt	None
Two	322-yards into Left Rough	Hit to 102-yards from the hole	On Green, 10-feet from hole	Makes 10-foot putt	None
Three	267-yards into Fairway	Hit into bunker beside Green	On Green, 16-feet from hole	Misses putt from 16-feet	Makes Putt
Four	176-yards into Rough	On Green, 2-yards from the hole	Makes 2-foot putt	None	None
Five	298-yards into Fairway	Hit to 150-yards from the hole	On Green, 21-feet from hole	Makes 21-foot putt	None

TABLE 1. Player A's Statistics

Now I will go through the statistics of these five holes. The **overall score average** was 20.00. Player A made two birdies, so he has a **birdie average** of 2.00. Player A's **driving distance** was 311-yards, since his two longest drives were 322 and 300-yards, and his **driving accuracy** was 75.00 since he hit three of the four fairways. Since Player A hit the green two strokes below par on three of the five holes, his **greens in regulation** stat is 60.00. Player A hit three greens in regulation and made two birdies when he did this, thus his **birdie or better conversion** is 66.67. Player A also missed two GIR's but was able to still make par on one of those holes, hence his **scrambling** is 50.00.

(8) *Strokes Gained Putting*: The number of putts a player takes from a specific distance is measured against a statistical baseline. The average number of putts a player takes from every distance is found, to determine the players strokes gained or lost on a hole. The sum of the values for all holes played in a round minus the field average strokes gained/lost for a round is the players strokes gained/lost for that round.

(9) *Strokes Gained Total*: The per round average of the number of strokes the player was better or worse than the field average on the same course and event.

(10) *Strokes Gained Tee-to-Green*: The per round average of the number of strokes the player was better of worse than the field average on the same course and event minus the Strokes Gained Putting.

I will now give an example to better describe the strokes gained statistics where I will still use the performance of Player A described above. Let's say that the PGA Tour average, or statistical baseline, for a putt of 30-feet, is 1.96 putts. This means from 30-feet, PGA Tour players take an average of 1.96 putts from this distance. Since Player A two-putted from this distance on Hole One, he loses 0.04 strokes, so at this point his strokes gained putting is -0.04. On the next four holes, his first putt on each was 10-feet, 16-feet, 2-feet, and 21-feet respectively. Let's assume that the PGA Tour average from these distances are 1.42, 1.75, 1.03, and 1.94 respectively. Since Player A, in order from these distances, putted once, twice, once and once, he gains 0.42, loses 0.25,

gains 0.03 and gains 0.94 strokes respectively. Thus his total strokes gained putting is $-0.04 + 0.42 - 0.25 + 0.03 + 0.94 = 1.1$ meaning Player A gained 1.1 strokes putting. Finally, this stat is taken for all players in this tournament and finds an average then that is subtracted from the individuals player statistics. If the field average for these five holes was 0.17, then Player A's **strokes gained putting** is 0.93. The strokes gained total takes the field average score and subtracts it from the individual players score. As stated, Player A shot a 20 on the five holes and let's assume that the field average was 21.19. This means that Player A's **strokes gained total** is 1.19. Finally, since strokes gained tee-to-green is the total subtracted by putting, Player A's **strokes gained tee-to-green** is 0.26.

All of these statistics will be used later when determining how a player can improve their scores.

3. Predicting Match Play

The game of golf has been explained through terms and examples and finally we are at the point where a winner of a match-play golf event will be predicted. In general, to solve this problem, two players on the PGA Tour will be pitted against each other and the odds of winning will be found. I know that this was possible based on the paper written by Gue, Smith, and Omen [**3**].

We have two players, Player A and Player B. Let X be the number of strokes that Player A uses at a single hole. Then we let $a_i = [P(X = i) \mid i = -2, -1, 0, 1, 2]$ where this is the probability of Player A shooting a specific score at a hole. We will define $i = -2$ as the event a player scores an *eagle* and $i = -1$ as the event a player scores a *birdie* on a hole. Let $i = 0$ be the event a player shoots a *par* on a hole. Also, let $i = 1$ be the event that a player scores a *bogey* and $i = 2$ be the event that a player shoots a *double bogey or worse* on a hole. The reason we combine *double bogey or worse* scores is because the odds of scoring worse than a *double bogey* on the PGA Tour are very rare and if a player scores this on a hole in match-play, they are almost guaranteed to lose the hole.

Similarly, let Y be the number of strokes taken on a hole for Player B and let $b_j = [P(Y = j) \mid j = -2, -1, 0, 1, 2]$ where j is defined in the same way as i.

On each hole there are three possibilities. The first possibility is Player A **wins** over Player B, which we will define as the probability that Player A wins a hole, denoted by

$$P(\text{A Win}) = \sum_{i=-2}^{1} \left(a_i \sum_{j=i+1}^{2} b_j \right).$$

The second possibility is Player A and Player B **tie** a hole, defined as the probability a tie occurs, denoted by

$$P(\text{Tie}) = \sum_{i=-2}^{2} \left(a_i \sum_{j=i}^{2} b_j \right).$$

The last possibility is Player A **loses** to Player B, and we will define this as the probability that Player A loses, denoted by

$$P(\text{A Lose}) = \sum_{i=-1}^{2} \left(a_i \sum_{j=-2}^{i-1} b_j \right).$$

These probabilities will be represented in a chart to show how likely one event is and all the entries will add up to one, since then an overall probability per hole of the three events can be found, which will look like:

	-2	-1	0	1	2
-2	Tie	A Win	A Win	A Win	A Win
-1	A Lose	Tie	A Win	A Win	A Win
0	A Lose	A Lose	Tie	A Win	A Win
1	A Lose	A Lose	A Lose	Tie	A Win
2	A Lose	A Lose	A Lose	A Lose	Tie

TABLE 2. a_i vs. b_j

As an example of how the above chart is filled, the space where both players shoot a birdie $(-1,-1)$ is $P(X=i,Y=j) = P(X=-1)P(Y=-1) = a_{-1}b_{-1}$. The space where Player A shoots a birdie and Player B shoots a par $(-1,0)$ hence Player A wins, is $P(X=i,Y=j) = P(X=-1)P(Y=0) = a_{-1}b_0$.

Each player will have a probability assigned to each score possible. This was found using the ESPN website to find the statistics of individual players [**5**]. The amount of times a specific score was shot was divided by the number of holes that player played in the 2014 season. In the following table, the statistics of four players are provided. The table shows their World Rank, holes played, amount of times each score was made, and the probability that the player shot that score on a hole, which is provided in parentheses besides the number of times each score was made.

World Rank	Player	Holes Played	Eagles	Birdies	Pars	Bogeys	Double Bogeys +
1	Rory McIlroy	1152	9 (0.0078)	293 (0.2543)	686 (0.5955)	141 (0.1224)	23 (0.0200)
9	Jordan Spieth	1764	4 (0.0023)	389 (0.2205)	1084 (0.6145)	263 (0.1491)	24 (0.0136)
10	Rickie Fowler	1530	4 (0.0026)	330 (0.2157)	964 (0.6301)	196 (0.1281)	36 (0.0235)
57	Graham DeLaet	1422	6 (0.0042)	298 (0.2096)	908 (0.6385)	184 (0.1294)	26 (0.0183)

TABLE 3. Player Scores

Then I calculated the probabilities of each event and the odds of a player winning, tying, or losing a hole to another player if they were to go head-to-head in a match-play event. One match I calculated was Rory McIlroy (Player A) versus Graham DeLaet (Player B) with the probabilities found shown below.

	-2	-1	0	1	2
-2	3.276×10^{-5}	1.635×10^{-3}	4.980×10^{-3}	1.009×10^{-3}	1.427×10^{-4}
-1	1.068×10^{-3}	5.330×10^{-2}	1.624×10^{-1}	3.291×10^{-2}	4.654×10^{-3}
0	2.501×10^{-3}	1.248×10^{-1}	3.802×10^{-1}	7.706×10^{-2}	1.090×10^{-2}
1	5.141×10^{-4}	2.566×10^{-2}	7.815×10^{-2}	1.584×10^{-2}	2.240×10^{-3}
2	8.400×10^{-5}	4.192×10^{-3}	1.277×10^{-2}	2.588×10^{-3}	3.660×10^{-4}

TABLE 4. Rory McIlroy (a_i) vs. Graham DeLaet (b_j)

As an example, the odds of tying a hole with a par is $(a_0 = 0.5955) \times (b_0 = 0.6385) = a_0 b_0 = 0.38024 = 3.802 \times 10^{-1}$. All the times a tie was possible, $a_i = b_j$, were added together to determine the probability of a tied hole. Similarly, when Rory McIlroy would win, $a_i < b_j$, were added together and when he lost, $a_i > b_j$. Below are the probabilities found for a player winning, tying, or losing a hole of match-play versus another player.

Match	A Wins	Tie	A Loses
Rory McIlroy (A) vs. Graham DeLaet (B)	0.29793 (29.793%)	0.44974 (44.974%)	0.25233 (25.233%)
Jordan Spieth (A) vs. Rickie Fowler (B)	0.27133 (27.133%)	0.45417 (45.417%)	0.27450 (27.450%)

TABLE 5. Event Probabilities

Now that we have the probability of one hole, we are finally ready to use our knowledge of Markov Chains to find the probability of winning a match-play event, where 18-holes of golf are played.

First thing we need to do is make sure that we are going to satisfy Definition 2.1.7. [**1**] which states that a *discrete-time Markov Chain* is a stochastic process satisfying:

(1) Discrete-Time
(2) Countable or finite state space
(3) Markov Property

Our process satisfies 1. a discrete-time stochastic process since the collection of random variables, $a_i = P(X = i)$ and $b_j = (Y = j)$, is defined by a set of possible outcomes; P(A Win), P (Tie), P (A Lose) making it stochastic. Furthermore, since the collection of random variables is countable, i.e. $i = j = -2, -1, 0, 1, 2$, we have a discrete-time process.

2. A countable or finite state space is assumed since we have a stochastic process and it was stated that is finite, thus the process is further classified as a *chain*.

Lastly, 3. the *Markov Property*, is satisfied because the future location depends on the present, since in order for a golfer to go **'2-up'**, they must win a hole after already being **'1-up'** or lose a hole after being **'3-up'**. This also suggests that going from one state to another is independent of the time in which it is being made, thus we have what we wanted, a *discrete-time stationary Markov Chain*.

We are ready to construct a matrix that will provide the probabilities of winning, tying, and losing after an 18-hole match-play event. Recall that we need a *transition matrix*, and Proposition 2.2.2. [**1**] tells us that a transition matrix must satisfy:

(1) All entries are non-negative
(2) The sum of the entries in each row is one

This is why it is important that at every hole it is true that,

$$P\,(\text{A Win}) + P\,(\text{Tie}) + P\,(\text{A Lose}) = 1.$$

This means that in our transition matrix, a vector can be placed in it at each row to add up to 1 and satisfy the properties of being a transition matrix. For convenience, we will call the transition matrix, P.

To start P, we need to determine where the chain starts. Since a match will start at the first hole the players are **'even'**. Furthermore, the match has not started and thus the starting vector, denoted $\mathbf{a}_0 = 1$, corresponds to the distribution at time zero. Then every vector after this will be of the form

$$\begin{bmatrix} l & t & w \end{bmatrix}$$

where $w = P\,(\text{A Win})$, $t = P\,(\text{Tie})$, $l = P\,(\text{A Lose})$. This vector, as stated, adds up to one and the t coordinate will be the point on the diagonal with the starting vector 1. This will continue until the last coordinate which will also be a 1 to signify the end of the match. Now we need to determine how large to make P. Since the maximum score a player can achieve in match-play is **'10-up'** (or **'10-down'** for the other player), in 10-holes, we only need to make the matrix 21×21 because we can go **'10-up'** to **'10-down'** with a tie in the middle. Remember from Chapter 3 Section 1 that this is possible because at this score a **dormie** will have been reached, making it impossible to come back and tie the match. Based on this description, matrix P will look like

$$P = \begin{bmatrix}
1 & 0 & 0 & 0 & ... & ... & ... & ... & 0 & 0 \\
l & t & w & 0 & ... & ... & ... & ... & 0 & 0 \\
0 & l & t & w & ... & ... & ... & ... & 0 & 0 \\
... & 0 & 0 & ... & ... & ... & ... & ... & ... & ... \\
... & 0 & 0 & ... & ... & 0 & ... & ... & .... & ... \\
... & ... & ... & ... & l & t & w. & ... & ... & ... \\
... & ... & ... & ... & ... & 0 & ... & 0 & 0 & ... \\
... & ... & ... & ... & ... & ... & l & t & w & 0 \\
0 & 0 & 0 & ... & ... & ... & ... & l & t & w \\
0 & 0 & 0 & ... & ... & ... & ... & 0 & 0 & 1
\end{bmatrix}_{21\times 21}$$

To give a more visual representation of this matrix, the digraph looks like

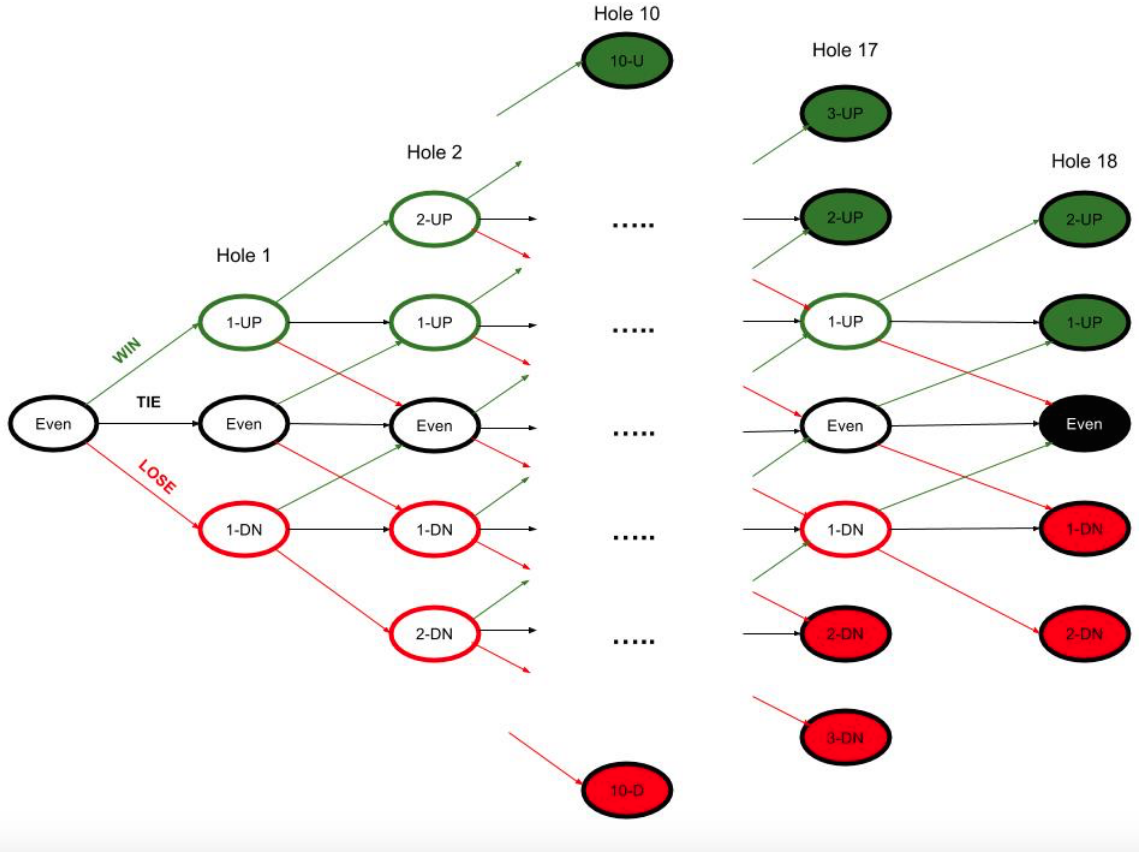

Figure 5

We will have P^{18} to signify the event of 18-holes of golf being played.

We are only interested in the probability of who will win, not by how much, thus P will be multiplied by a 21×3 matrix to signify the event of winning, tying, or losing. By doing this, an overall probability of winning, tying, and losing, instead of the individual probabilities of these events will be achieved. This means that we will receive $P\,(\text{A Win})$, not $P\,(\text{A Win '10-up'})$ and $P\,(\text{A Win '9-up'})$ and so on. The first row will be the event A wins, second row will be the event of a tie and the last row will be the event A Loses. This means that the first ten entires in row one, will have a 1 since this signifies $P\,(\text{A Win})$. Similarly, the last ten entries in the third row will have 1 and the middle entry in the second row will have a 1. All other entries will be 0 and the matrix, say R, will look like:

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^{T}_{3\times 21}$$

It is also necessary to multiply P by a 1×21 vector, v, which will just have an entry of 1 in the middle (or 11^{th} entry) and 0's elsewhere. This will give an overall probability of the three outcomes adding to one.

We are now ready to compute the results of 18-holes of match-play where the results from Table 5 will be put into P. Once we have computed our desired matches, we will have the probability that Player A wins, ties, or loses to Player B after an 18-hole match-play event. Recall Rory McIlroy (A) vs. Graham DeLaet (B) by letting

$$\begin{aligned} w &= \text{A Wins} = 0.29793 \\ t &= \text{Tie} = 0.44974 \\ l &= \text{A Loses} = 0.25233 \end{aligned}$$

and we compute $vP^{18}R$ for this match and the other one described in Table 5 and we get the following results:

Match	A Wins	Tie	A Loses
Rory McIlroy (A) vs. Graham DeLaet (B)	0.54164 (54.164%)	0.12161 (12.161%)	0.33675 (33.675%)
Jordan Spieth (A) vs. Rickie Fowler (B)	0.42967 (42.967%)	0.12622 (12.622%)	0.44411 (44.411%)

TABLE 6. 18-Hole Match-Play Probabilities

These results yield some interesting facts. Even though the odds of tying one hole are very good, the odds of tying a match are much lower. As well, Rory McIlroy was found after one hole to be favoured by a small margin and after 18-holes, the odds are highly in his favour. The match between Jordan Spieth and Rickie Fowler shows such a slight variation between the two, quite similar to that after one hole. As expected though, Rory McIlroy should win over Graham DeLaet and Rickie Fowler has a slight advantage over Jordan Spieth.

The reason why it is important to use the results after 18-holes is because the results can vary greatly. Yes, there was a high chance of a tied hole but the odds of a match being tied is significantly lower. Consider theoretically, two golfers very evenly matched, such that

$$\begin{aligned} P(\text{A Win}) &= 0.09 \\ P(\text{Tie}) &= 0.85 \\ P(\text{A Lose}) &= 0.06 \end{aligned}$$

after one hole, but after 18-holes we see that

$$\begin{aligned} P(\text{A Win}) &= 0.50402 \\ P(\text{Tie}) &= 0.24020 \\ P(\text{A Lose}) &= 0.25578 \end{aligned}$$

There is a 50% chance that Player A wins the match even though his odds of winning a hole are just 9% and the odds of tying the match are massively reduced. The variation

after 18-holes to one hole of match-play alters significantly, thus the entire round must be taken into effect in order to get a true understanding of the probability of being victorious in a match-play event.

4. Stroke-Play

In this section, we will use the knowledge of Markov Chains to extend to predicting stroke-play events. This allows us to see if there is a change in the odds of winning, tying and losing when compared to match-play, along with other predictions to see if there is an advantage to playing the different styles. In order to do this, the data from **Table 3. Player Scores** will be used. In Table 3, the probability of certain scores a player shot in 2014 were found [**5**]. The amount of times a score occurred was divided by the holes played. As an example, the odds for Rory McIlroy are displayed with notation:

$$\begin{aligned} \text{Double Bogeys + (D)} &= 0.0200 \\ \text{Bogeys (O)} &= 0.1224 \\ \text{Pars (P)} &= 0.5955 \\ \text{Birdies (B)} &= 0.2543 \\ \text{Eagles (E)} &= 0.0078 \end{aligned}$$

The transition matrix, P, in this case will be slightly different than that used for match-play. The dimension of P will stay 21×21 since the odds of scoring a **'10-over'** or **'10-under'** round rarely happen and even worse or better than this is more unlikely. Starting again with the zero-vector 1 and ending with a 1 as well due to the same reasons as before. The entries used will be the score above so the matrix will consist of a 1×5 vector instead of a 1×3 used in match-play. However, the second entry will combine the odds of a double bogey and bogey and the second last entry will combine the odds of a birdie and eagle in order for the matrix to be computed. Based on this, P will look like:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & ... & ... & ... & ... & 0 & 0 \\ E+B & P & O & D & 0 & ... & ... & ... & 0 & 0 \\ E & B & P & O & D & 0 & ... & ... & 0 & 0 \\ 0 & E & B & P & O & D & ... & ... & ... & ... \\ ... & 0 & ... & ... & ... & ... & ... & ... & .... & ... \\ ... & ... & 0 & E & B & P & O & D & 0 & ... \\ ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & E & B & P & O & D \\ 0 & 0 & 0 & ... & ... & ... & E & B & P & D+O \\ 0 & 0 & 0 & ... & ... & ... & ... & 0 & 0 & 1 \end{bmatrix}_{21\times 21}$$

As with match-play we will compute P^{18} since there are 18-holes. This will be multiplied by a 1×21 vector, v, described before. Once again, using the statistics in Table 3, vP^{18} is computed and we see that:

Score	Rory McIlroy	Jordan Spieth	Rickie Fowler	Graham DeLaet
10	1.5371×10^{-4}	2.1854×10^{-4}	4.0224×10^{-4}	2.3303×10^{-4}
9	2.0233×10^{-4}	3.5631×10^{-4}	5.3335×10^{-4}	3.5570×10^{-4}
8	6.4524×10^{-4}	1.1453×10^{-3}	1.5736×10^{-3}	1.1181×10^{-3}
7	1.7044×10^{-3}	3.0676×10^{-3}	3.8818×10^{-3}	2.9436×10^{-3}
6	4.0822×10^{-3}	7.3622×10^{-3}	8.6661×10^{-3}	6.9941×10^{-3}
5	8.9420×10^{-3}	1.5890×10^{-2}	1.7639×10^{-2}	1.5053×10^{-2}
4	1.7882×10^{-2}	3.0760×10^{-2}	3.2631×10^{-2}	2.9240×10^{-2}
3	3.2539×10^{-2}	5.3213×10^{-2}	5.4617×10^{-2}	5.1011×10^{-2}
2	5.3668×10^{-2}	8.1959×10^{-2}	8.2293×10^{-2}	7.9508×10^{-2}
1	7.9905×10^{-2}	1.1196×10^{-1}	1.1101×10^{-1}	1.1011×10^{-1}
0	1.0695×10^{-1}	1.3515×10^{-1}	1.3329×10^{-1}	1.3473×10^{-1}
-1	1.2813×10^{-1}	1.4366×10^{-1}	1.4164×10^{-1}	1.4493×10^{-1}
-2	1.3686×10^{-1}	1.3402×10^{-1}	1.3249×10^{-1}	1.3641×10^{-1}
-3	1.2983×10^{-1}	1.0941×10^{-1}	1.0855×10^{-1}	1.1192×10^{-1}
-4	1.0903×10^{-1}	7.7957×10^{-2}	7.7556×10^{-2}	7.9840×10^{-2}
-5	8.0818×10^{-2}	4.8362×10^{-2}	4.8154×10^{-2}	4.9419×10^{-2}
-6	5.2741×10^{-2}	2.6059×10^{-2}	2.5903×10^{-2}	2.6510×10^{-2}
-7	3.0223×10^{-2}	1.2163×10^{-2}	1.2037×10^{-2}	1.2311×10^{-2}
-8	1.5094×10^{-2}	4.8879×10^{-3}	4.8018×10^{-3}	4.9294×10^{-3}
-9	6.2812×10^{-3}	1.6179×10^{-3}	1.5858×10^{-3}	1.6397×10^{-3}
-10	4.3280×10^{-3}	7.8234×10^{-4}	7.4747×10^{-4}	7.9426×10^{-4}

TABLE 7. Scores Relative to Par Probabilities

If the data for Rory McIlroy and Graham DeLaet is plotted on a graph, we see:

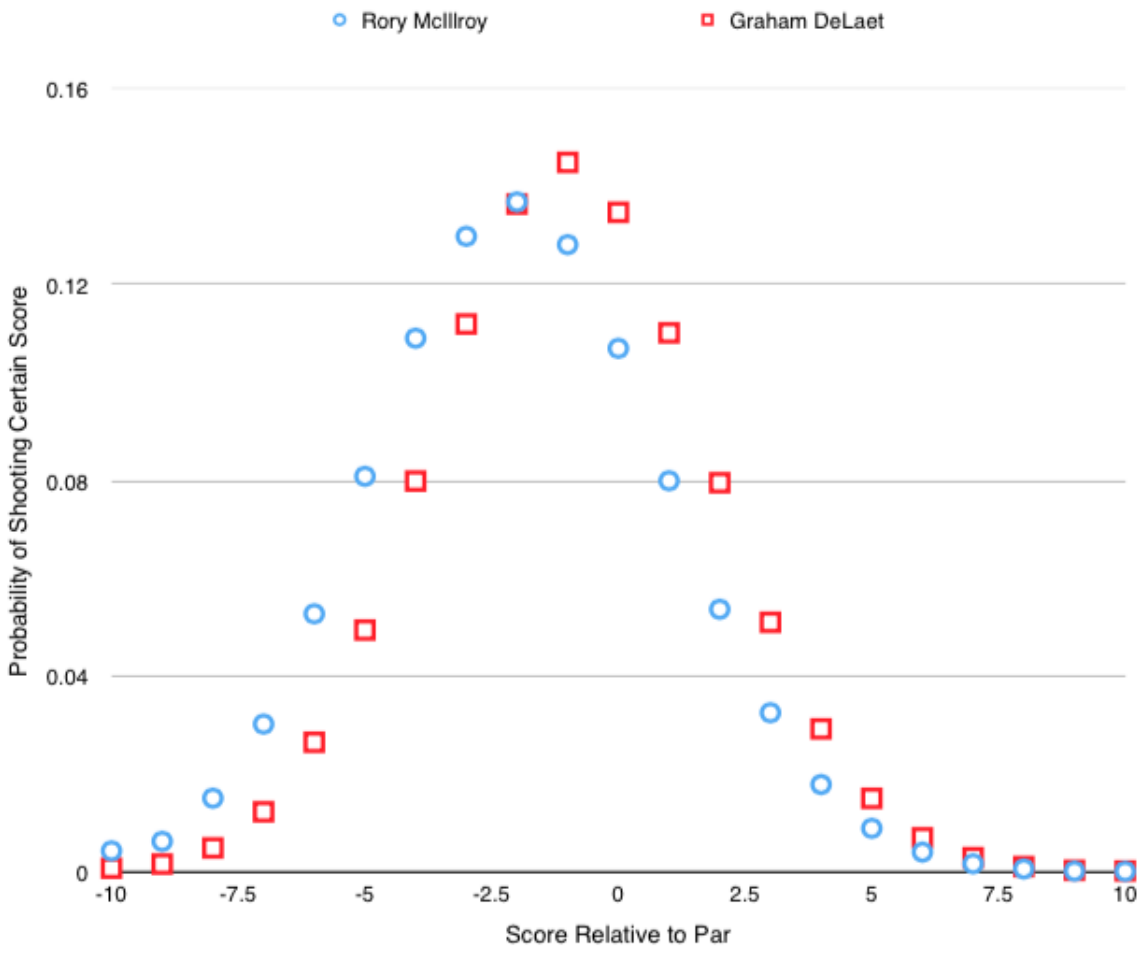

Figure 6

Now we want to determine the probability of a player beating another in stroke-play. Similar to that of match-play, we will take the data from the Markov Chain and combine the odds similar to one hole of match-play, except going backwards. We want the odds of winning after one round, which is why this is done. We have Player A and Player B. Let X be the number of strokes relative to the par of a golf course that Player A uses in a round. We let $a_i = [P(X = i) \mid i = -10, -9, ..., -1, 0, 1, ..., 9, 10]$ where this is the probability of Player A shooting a specific score of a round. Define $i = -10$ as a **'10'-under'** par round, $i = -9$ as a **'9-under'** par round and so on to $i = 10$ as a **'10-over'** par round of golf. Similarly, Y represents the strokes taken by Player B over a round and $b_j = [P(Y = j) \mid j = -10, -9, ..., -1, 0, 1, ..., 9, 10]$.

Each round has three possibilities. The first is Player A shooting a lower round than Player B, hence **winning**, denoted by

$$P(\text{A Win}) = \sum_{i=-10}^{9} \left(a_i \sum_{j=i+1}^{10} b_j \right).$$

Player A and Player B shooting the same score, hence **tying**, is denoted by

$$P(\text{Tie}) = \sum_{i=-10}^{10} \left(a_i \sum_{j=i}^{10} b_j \right).$$

Finally, Player A shooting a higher round than Player B, hence **losing**, is denoted by

$$P(\text{A Lose}) = \sum_{i=-9}^{10} \left(a_i \sum_{j=-10}^{i-1} b_j \right).$$

Recall Table 2. a_i vs. b_j. The probabilities will be calculated in the exact same way as Table 2 except both the row and column will be from -10 to 10. Then, the probabilities were calculated and put into the corresponding spot, similar to Table 4. Rory McIlroy (a_i) vs. Graham DeLaet (b_j). Once this was done, the probabilities of winning, tying, and losing in a stroke-play event were found, using the same matches as match-play:

Match	A Wins	Tie	A Loses
Rory McIlroy (A) vs. Graham DeLaet (B)	0.5499 (54.99%)	0.0962 (9.62%)	0.3539 (35.39%)
Jordan Spieth (A) vs. Rickie Fowler (B)	0.4540 (45.40%)	0.1006 (10.06%)	0.4454 (44.54%)

TABLE 8. Probabilities of Stroke-Play

Recall Table 6. 18-Hole Match-Play Probabilities to compare the results of stroke-play vs. match-play of these matches.

When comparing stroke-play to match-play some interesting results are found. First, notice the probability of tying in match-play is greater than that of stroke-play. This makes sense, since in stroke-play there are five outcomes on each hole instead of three, which makes a tie less likely to occur.

Secondly, as with match-play, Rory McIlroy is expected to do better than Graham DeLaet, even though Graham has improved his odds. If we look at the graph in Figure 6. this does make sense since Rory McIlroy has a slightly lower standard deviation. This is directly related to how well a player will score. The lower the standard deviation, the more consistent a player is. In stroke-play this is important because the more consistent a player is, the less likely they are to shoot high scores on holes. Since stroke-play is an accumulation of strokes, the lower the scores the better, and being more consistent to avoid high scores is a direct testament of this.

Lastly, and most fascinating, are the results of Jordan Spieth vs. Rickie Fowler. Recall that in match-play, Jordan Spieth had about a 43% and 44% chance of winning and losing respectively, so in match-play he is more likely to lose. Comparing these results with stroke-play, Jordan Spieth has about a 45% and 44% chance of winning and losing respectively; hence, he is now more likely to win! This is exciting since it tells us that there is a difference in outcomes between the two playing styles. Just like with the other match, after looking at the standard deviations of Jordan Spieth and Rickie Fowler, it is found that Jordan has a lower standard deviation, thus he is more consistent. This helps explain why we have a difference in who will win or lose between these two golfers. Since Jordan Spieth is more consistent, he is less likely to shoot really high scores on a hole, so in stroke-play he keeps his score to a minimum. Having a bad hole in match-play is not as big a deal since you would just lose that hole and then restart at the next. Since stroke-play is the accumulation of strokes, having a bad hole affects the round and one has to make up these shots. By being more consistent, a player will not have as many bad holes and keep their score to a minimum, which doesn't necessarily pay off in match-play. Consider a double bogey (**+2**) by one player and a par (**even**) by the other on the same hole. In match-play, this is just one lost hole, or won, but in stroke-play this is a two shot swing. Thus being a more consistent player allows you to gain more strokes on a hole proving to be beneficial in stroke-play and the reason why Jordan Spieth is more likely to beat Rickie Fowler in a stroke-play event but more likely to lose in a match-play event.

CHAPTER 4

Further Application

1. Correlation

In this chapter, a different branch of mathematics, called **correlation**, will be used to determine the best way a player can improve their game on the PGA Tour. This is of interest because if a player improves their game, they will obviously become better and have a greater chance of winning match and stroke-play events. This chapter will use the statistics described in Chapter 3: Section 2 to find the best way a player can improve. The information in this chapter was taken from [**2**], except for the statistics of the PGA Tour players.

Definition 4.1: *Correlation* refers to the relationship between two random variables or two data sets involving dependence.

In a correlation model there are two random variables, say X and Y. Correlation measures the strength of a linear relationship that exists between X and Y. In order to determine how strong a relationship is, it is mandatory to use the data points and find an estimator for X and Y.

Definition 4.2: The theoretical parameter used to measure the linear relationship between X and Y is the *Pearson correlation coefficient*, ρ, defined by

$$\rho = \frac{Cov(X,Y)}{\sqrt{(VarX)(VarY)}}.$$

Since it is of interest to find just an estimate, the *estimator* for ρ is given by

$$\hat{\rho} = R = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}.$$

Theorem 4.3: The correlation coefficient ρ for any two random variables X and Y lies between -1 and 1.

In order to calculate an estimator for the Pearson correlation coefficient, all the variables within $\hat{\rho}$ must be calculated. These variables are based on a set $\{(X_i,\, Y_i) : i = 1, 2, 3, ..., n\}$

of n observations on the random variable (X, Y) and they have the following equations

$$\begin{aligned} S_{xx} &= n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2 \\ S_{yy} &= n\sum_{i=1}^{n} y^2 - \left(\sum_{i=1}^{n} y\right)^2 \\ S_{xy} &= n\sum_{i=1}^{n} xy - \left(\sum_{i=1}^{n} x \sum_{i=1}^{n} y\right). \end{aligned}$$

Statistics from 20 players on the PGA Tour of various World Rankings were found on the PGA Tour website from the 2014 season [**8**], and shown below:

World Ranking	Player	Stroke Average	Birdie Average	Driving Distance	Driving Accuracy	Greens in Regulation	Birdie/Better Conversion(%)	Scrambling(%)	StrokesGained: Putting	StrokesGained: Tee-to Green	StrokesGained: Total
1	Rory McIlroy	68.827	4.58	310.5	59.93	69.44	36.84	58.52	0.273	1.993	2.266
3	Adam Scott	69.205	4.24	303.5	61.51	68.79	33.89	60.73	0.217	1.503	1.720
4	Bubba Watson	69.712	3.96	314.3	60.47	67.96	32.40	59.14	-0.053	1.322	1.269
5	Sergio Garcia	68.950	3.86	294.3	62.19	68.68	29.97	66.67	0.164	1.820	1.984
8	Jason Day	69.626	3.58	301.2	58.72	64.00	30.02	64.95	0.316	0.849	1.165
9	Jordan Spieth	69.946	3.97	289.7	58.79	62.47	34.24	62.39	0.398	0.512	0.910
10	Rickie Fowler	70.172	3.88	297.5	59.93	64.51	32.52	62.06	0.287	0.419	0.706
12	Martin Kaymer	70.468	3.69	294.9	61.27	64.58	31.00	53.43	-0.114	0.358	0.244
13	Billy Horchel	70.521	3.59	291.6	67.04	70.43	28.41	55.32	0.226	0.451	0.677
14	Phil Mickelson	70.278	3.75	292.4	58.01	65.26	31.29	61.71	0.249	0.471	0.720
21	Jimmy Walker	69.789	4.18	301.0	51.67	67.59	33.50	58.39	0.482	0.681	1.163
23	Patrick Reed	70.598	3.49	292.3	55.59	63.64	30.25	59.42	0.245	0.061	0.306
33	Luke Donald	70.735	3.43	278.1	62.56	63.00	29.18	60.59	0.520	0.170	0.690
43	Webb Simpson	70.234	3.75	288.5	62.24	65.28	30.98	59.05	0.301	0.575	0.876
57	Graham DeLaet	69.943	3.77	303.4	62.07	70.68	28.91	58.75	-0.095	1.191	1.096
63	Ernie Els	70.736	3.28	291.2	53.26	61.60	29.24	61.36	-0.176	0.659	0.483
73	Angel Cabrera	70.675	3.68	303.7	58.40	64.98	30.06	53.10	0.011	0.194	0.205
85	Cameron Tringale	70.630	3.35	282.9	63.87	66.77	27.03	59.34	0.111	0.490	0.601
94	Geoff Ogilvy	71.422	3.30	292.5	58.51	64.51	28.05	55.77	-0.333	-0.114	-0.447
100	Russell Knox	70.111	3.39	285.7	67.60	68.00	27.11	63.40	-0.120	0.860	0.740

TABLE 1. PGA Tour Player Statistics

Now, in order to determine the best way of improving a players score, *Stroke Average* is correlated against other statistics. This is to see if there is a strong correlation between the stroke average, where the lower the better, and other statistics. If a strong correlation is exhibited, then improving this statistic would benefit players on the PGA Tour. Of course most of the statistics if improved would be beneficial to the stroke average but this is to see what statistics are better than others.

To achieve the overall goal, every time the stroke average is correlated with a different statistic, we let X = Stroke Average, Y = Other Statistic (such as scrambling %), and $n = 20$, since we are using the statistics from 20 players. As an example, stroke average

is always represented by S_{xx} and this equation is used and yields the following;

$$\begin{aligned} S_{xx} &= n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2 \\ &= 20\left(68.827^2 + 69.205^2 + 69.712^2 + 68.950^2 + ... + 71.422^2 + 70.111^2\right) \\ &\quad - \left(68.827 + 69.205 + 69.712 + 68.950 + ... + 71.422 + 70.111\right)^2 \\ &= 20(98.368.776) - (1402.578)^2 \\ S_{xx} &= 150.482 \end{aligned}$$

S_{yy} is done in the same manner and S_{xy} will also be similar, except each X_i will be multiplied by Y_i then subtracted from the sum of X multiplied with Y.

The data for X and Y was imputed into a computer and the correlation between two statistics was found

X	Y	$\hat{\rho}$
Stroke Average	Birdie Average	-0.80058
Stroke Average	Driving Distance	-0.55269
Stroke Average	Driving Accuracy	-0.06709
Stroke Average	Greens in Regulation	-0.52619
Stroke Average	Birdie/Better Conversion %	-0.63340
Stroke Average	Scrambling %	-0.48325
Stroke Average	Strokes Gained: Putting	-0.35449
Stroke Average	Strokes Gained: Tee-to-Green	-0.91546
Stroke Average	Strokes Gained: Total	-0.96421

TABLE 2. Correlation Results

To better grasp the results achieved, I will look at the significance level. For correlation with a data set of $n = 20$, we use $n - 2 = 20 - 2 = 18$ to determine the significance level. If $|\hat{\rho}| = 0.444$ or higher, there is a 0.05 significance level, or 95% confidence, and if $|\hat{\rho}| = 0.561$ or higher then there is a 0.01 significance level, or 99% confidence [**2**]. Based on our results, everything but *driving accuracy* and *strokes gained putting* had at least a 0.05 significance level. This tells us that improving these aspects of the game will not significantly improve a players game. This is extremely surprising since it is advised to improve these aspects of the game on the PGA Tour! Yes there are many different aspects to consider and not everyone is the same, but based on the data with these 20 players, shooting lower scores can be better achieved through improving other statistics. Every other other statistic has a correlation of 0.05 significance level or better.

In conclusion, lowering ones' score can be achieved by improving any of the statistics described but despite popular belief, improving driving accuracy will not greatly improve a players score, thus changing the results found in Chapter 3.

CHAPTER 5

Concluding Remarks

Markov Chains are a very useful tool in mathematics and can be extended to various applications in the world to predict certain outcomes. Strictly based on this paper, a very powerful application to Markov Chains is of course sports, more specifically, golf. Golf is a wonderful sport played by millions both at the professional level and by the average person looking to go out and have some fun. This paper looked at discrete-time Markov Chains but there is a word of knowledge and extensions that can be made to this type of Markov Chain, as well as continuos-time Markov Chains.

In this paper, we showed many different properties that define a Markov Chain which was used in determining the results of match-play and then stroke-play golf events. We showed how to get the odds of one golfer beating another in a match-play event by looking at the odds of a certain score that a player shoots on a hole. This knowledge was then extended to finding the odds of a player shooting a specific score in stroke-play. Then we determined how the matches could end differently based on the two different playing events. It was in fact found that a difference between the two events can occur due to the overall consistency of a player. Finally, using correlation, the best way to improve a players' score on the PGA Tour was found, with an interesting result, stating that driving accuracy was the least significant way to lower ones' score and shoot better rounds of golf.

In the future, it would be interesting to continue the study of Markov Chains when applied to golf, in particular, finding a different way or method to determine the probabilities of shooting specific scores. Instead of taking the total odds of a score on a hole, the odds could be split between the par 3's, 4's, and 5's to possibly get a more accurate result, as the par of a hole is affected and played differently by each player. It would also be of interest to use the statistics described in Chapter 3, Section 2 and come up with some way to use these in order to find the odds of shooting specific scores, as opposed to just using the odds of shooting a certain score on a hole as was done in this paper.

In conclusion, this paper just touched the surface of Markov Chains. Many more applications and outcomes can be found and studied using the incredible results we can achieve through Markov Chains.

Bibliography

[1] D.L. Isaacson and R.W. Madsen. *Markov Chains: Theory and Applications*. John Wiley and Sons, Inc. New York, New York. (1976). 1-22, 43-60 2, 8, 20, 21

[2] J.C. Arnold and J.S. Milton. *Introduction to Probability and Statistics 4th Edition*. McGraw-Hill Inc. New York, New York. (2003). 418-423 29, 31

[3] K.R. Gue, J. Smith, O. Ozmen. *Predicting Results of a Match-Play Golf Tournament with Markov Chains*. (2010) 18

[4] Lay, David. *Linear Algebra and Its Applications 4th Edition*. Pearson Education Inc. Boston, Massachusetts. (2012). 253-257 2, 7, 8

[5] http://espn.go.com/golf/schedule/$_{/}year/2014.2015.ESPN, Inc.Web.19, 24$

[6] http://www.igfgolf.org/about-golf/history/. 2014. International Golf Federation. Web. 13

[7] http://www.pga.com/golf-instruction/instruction-feature/fundamentals/golf-glossary-and-golf-terms. 2014. PGA Tour, Inc. Web. 13, 14, 15

[8] http://www.pgatour.com/players.html. 2014. PGA Tour, Inc. Web. 15, 16, 30