Markov Chains and its Applications to Golf

Mason Weld

April 20, 2018

1 Match-Play Golf

1.1 Rules and Scoring

Match-play golf is a competition between two golfers. The scoring is determined by comparing the amount of strokes that it takes both golfer A and golfer B to complete the hole. We can denote the amount of strokes it takes golfer A and B to complete the ith hole as A_i and B_i respectively, for $i = 1, 2, 3, ..., 18$. If we approach the scoring from golfer A's perspective, then $A_i < B_i$ means golfer A has won the hole, while $A_i > B_i$ means golfer A has lost the hole. If $A_i = B_i$, then golfer A has tied the hole. At the beginning of the match (before any holes have been played), the match is said to be 'all square'. All square describes a match that is currently tied. After the completion of the first hole, if $A_1 = 4$ and $B_1 = 5$, looking from golfer A's perspective, the match is now at '1-up'. Conversely, golfer B is said to be '1-down'. If both golfers take the same amount of strokes on hole number 2, the score of the match stays the same (golfer A is still '1-up'). In general, if golfer A wins n holes, and loses m holes, the score in golfer A's perspective is $n-m$. A positive integer score indicates the amount golfer A is 'up' in the match, while a negative integer score indicates the amount golfer A is 'down' in the match. If $n - m = 0$, then the match is at 'all square'. The match is over when either 18 holes are completed, or when a player has no chance of winning. For example, if golfer A is '4-down' with only 3 holes left out of the total 18, then the match will end with golfer B being victorious because golfer A does not have enough holes to comeback from the '4-down' deficit. The maximum score a golfer can achieve in match-play is '10-up'. This is because a golfer does not have enough holes to comeback from a '10-down' deficit.

1.2 Hole Probabilities

There are three possible outcomes for each hole in match-play. Either golfer A wins, ties, or loses the hole. The sum of the probabilities of these three events has to equal 1, according to the second axiom of probability. This can be denoted as:

$$P(A^{win}) + P(A^{tie}) + P(A^{lose}) = 1 \quad (1)$$

To calculate these probabilities for each golfer, we will denote α_x as the probability that golfer A takes x strokes on the hole being evaluated. Similarly, the probability that golfer B takes y strokes on the specific hole is β_y. Since PGA professional golfers rarely ever take more than 8 strokes for a given hole, we will only consider $x = 1, 2, ..., 8$ and $y = 1, 2, ..., 8$. It follows that the probability that golfer A tying the hole is:

$$P(A^{tie}) = \sum_{k=1}^{8} \alpha_k \beta_k \quad (2)$$

The probability of golfer A winning the hole is:

$$P(A^{win}) = \sum_{x=1}^{7} \sum_{y=x+1}^{8} \alpha_x \beta_y \quad (3)$$

And finally, the probability of golfer A losing the hole is:

$$P(A^{lose}) = \sum_{x=2}^{8} \sum_{y=1}^{x-1} \alpha_x \beta_y \tag{4}$$

We can visualize these probabilities in a table.

	β_y							
α_x	1	2	3	4	5	6	7	8
1	$\alpha_1\beta_1$	$\alpha_1\beta_2$	$\alpha_1\beta_3$	$\alpha_1\beta_4$	$\alpha_1\beta_5$	$\alpha_1\beta_6$	$\alpha_1\beta_7$	$\alpha_1\beta_8$
2	$\alpha_2\beta_1$	$\alpha_2\beta_2$	$\alpha_2\beta_3$	$\alpha_2\beta_4$	$\alpha_2\beta_5$	$\alpha_2\beta_6$	$\alpha_2\beta_7$	$\alpha_2\beta_8$
3	$\alpha_3\beta_1$	$\alpha_3\beta_2$	$\alpha_3\beta_3$	$\alpha_3\beta_4$	$\alpha_3\beta_5$	$\alpha_3\beta_6$	$\alpha_3\beta_7$	$\alpha_3\beta_8$
4	$\alpha_4\beta_1$	$\alpha_4\beta_2$	$\alpha_4\beta_3$	$\alpha_4\beta_4$	$\alpha_4\beta_5$	$\alpha_4\beta_6$	$\alpha_4\beta_7$	$\alpha_4\beta_8$
5	$\alpha_5\beta_1$	$\alpha_5\beta_2$	$\alpha_5\beta_3$	$\alpha_5\beta_4$	$\alpha_5\beta_5$	$\alpha_5\beta_6$	$\alpha_5\beta_7$	$\alpha_5\beta_8$
6	$\alpha_6\beta_1$	$\alpha_6\beta_2$	$\alpha_6\beta_3$	$\alpha_6\beta_4$	$\alpha_6\beta_5$	$\alpha_6\beta_6$	$\alpha_6\beta_7$	$\alpha_6\beta_8$
7	$\alpha_7\beta_1$	$\alpha_7\beta_2$	$\alpha_7\beta_3$	$\alpha_7\beta_4$	$\alpha_7\beta_5$	$\alpha_7\beta_6$	$\alpha_7\beta_7$	$\alpha_7\beta_8$
8	$\alpha_8\beta_1$	$\alpha_8\beta_2$	$\alpha_8\beta_3$	$\alpha_8\beta_4$	$\alpha_8\beta_5$	$\alpha_8\beta_6$	$\alpha_8\beta_7$	$\alpha_8\beta_8$

Figure 1: Match-Play Probabilities

To interpret the table, take the example of golfer A taking 3 strokes and golfer B taking 4 strokes to complete the same hole. The probability of this event happening is $\alpha_3\beta_4$, which corresponds to the index $(3, 4)$ in the table. Another thing to note is that this index is colored green. That is because the color corresponds the scoring of that hole, in golfer A's perspective. A green entry corresponds to A winning the hole, yellow corresponds to a tie, and red corresponds to A losing the hole. Therefore, equation (2) can be visualized by the summation of all the yellow entries. For the same reason, equation (3) and (4) can be visualized by the summation of the green and red entries, respectively.

The most simple way of assigning values to α and β is to look at the season long statistics of players A and B.

	1 Shot Holes	2 Shot Holes	3 Shot Holes	4 Shot Holes	5 Shot Holes	6 Shot Holes	7 Shot Holes	8 Shot Holes	Total Holes
Rickie Fowler	0 (0.0000)	54 (0.0395)	364 (0.2661)	550 (0.4020)	364 (0.2661)	32 (0.0234)	4 (0.0029)	0 (0.0000)	1368
Dustin Johnson	0 (0.0000)	43 (0.0341)	324 (0.2571)	530 (0.4206)	331 (0.2627)	29 (.0230)	3 (0.0024)	0 (0.0000)	1260

Figure 2: Fowler and Johnson 2017 Statistics

The number in the parenthesis indicates the probability of the golfer completing a hole in the corresponding number of strokes. For example, if Rickie Fowler is golfer A, then $\alpha_4 = 0.4020$. Now we have all the data needed to fill out Figure 1 for Rickie Fowler (A) and Dustin Johnson (B).

α_x \ β_y	1	2	3	4	5	6	7	8
1	0	0	0	0	0	0	0	0
2	0	0.00135	0.01016	0.01661	0.01038	0.00091	0.0001	0
3	0	0.00907	0.06841	0.11192	0.0699	0.00612	0.00064	0
4	0	0.01371	0.1034	0.16908	0.10561	0.00925	0.00096	0
5	0	0.00907	0.06839	0.0951	0.0699	0.00612	0.00064	0
6	0	0.0008	0.00602	0.00984	0.00615	0.00054	0.00006	0
7	0	0.00099	0.00075	0.00122	0.00076	0.00007	0.00001	0
8	0	0	0	0	0	0	0	0

Figure 3: Fowler and Johnson Hole Probabilities

Summing all the entries of this table where $x < y$ (green colored entries) will give the probability of Fowler winning the hole $P(A^{win}) = 0.34938$. Using the same logic for entries where $x > y$, we obtain the probability of Fowler losing the hole $P(A^{lose}) = 0.32534$, and the probability of the hole being tied $P(A^{tie}) = 0.30929$. Now we are ready to use a Markov Model to predict the outcome of this match.

1.3 Simple Markov Model

A sequence of random variables, say $\{X_1, X_2, X_3, ...\}$, with state space $\{s_1, s_2, s_3, ...\}$, is considered a Markov Chain if

$$P(X_n = s_n | X_0 = s_0, X_1 = s_1, ..., X_{n-1} = s_{n-1}) = P(X_n = s_n | X_{n-1} = s_{n-1}) \tag{5}$$

This property is important, because it shows that the probability of moving to a specific state only depends on the current state, and does not depend on the previous states. This memory-less property is commonly referred to as the Markov property.

Before proving that our match-play golf model follows the Markov property, we will define our state space. Recall that '10-up' (or '10-down') is the highest (and lowest) score that can be achieved in the match. For the sake of notation, I will declare state -3 as golfer A being '3-down' in the match, while state 3 will represent golfer A being '3-up' in the match. Therefore, our model will have state space $S = \{-10, -9, ..., 0, ..., 9, 10\}$ and $|S| = 21$. Thinking about match-play golf, the probability of moving from state 0 (all-square) to state 1 (1-up) on the second hole is the same as it is on the sixteenth hole. More generally, the probability of moving from score n to k is the same for all 18 holes.

Now we must construct our transition matrix T. For consistency, we will always use right transition matrices. By the definition of a right transition matrix, the matrix must be square, and the sum of each row must be 1. Our transition matrix $T \in R^{21*21}$, is square because we have 21 states for the match. Recall from equation (1), that the sum of all the probabilities of the three outcomes is 1. Therefore, when we construct each row to have all zeros except for entries $P(A^{win}), P(A^{tie})$, and $P(A^{lose})$, we can be sure each row sums to 1.

$$
T = \begin{array}{c c}
 & \begin{array}{ccccccccc} -10 & -9 & -8 & -7 & & 7 & 8 & 9 & 10 \end{array} \\
\begin{array}{c} -10 \\ -9 \\ -8 \\ \\ \\ \\ 8 \\ 9 \\ 10 \end{array} &
\left[\begin{array}{ccccccccc}
1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\
P(A^{lose}) & P(A^{tie}) & P(A^{win}) & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & P(A^{lose}) & P(A^{tie}) & P(A^{win}) & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & P(A^{lose}) & P(A^{tie}) & P(A^{win}) & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 & P(A^{lose}) & P(A^{tie}) & P(A^{win}) \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}
$$

Figure 4: Initial Transition Matrix

Note that indices (1,1) and (21,21) have 1 as their entries. This is called an absorbing state. This means that once the Markov Chain reaches state -10 or 10, it will be return to the same state in the next step, making it impossible to ever leave this state once it is reached. In a golf match-play perspective, this makes sense because once state -10 or 10 is reached, the match has been won or lost by golfer A (refer back to the scoring rules of match-play). Following the same logic, after hole 11, states -9 and 9 must become absorbing states because the deficit is too large for the loser to make a comeback. In general, after $20 - l$ holes, states $-l$ and l become absorbing states for $l = 2, 3, ..., 9$. Therefore, for the first 10 holes, we will use the initial transition matrix T shown in Figure 4. After hole 10, we will adjust the transition matrix by adding the required absorbing states for each hole until hole 18. The transition matrix used for hole n (after hole 10) will be denoted T_n for $n = 11, 12, ..., 18$.

The last thing we need to for our Markov Model, is create our stochastic probability vector $\mathbf{x}$. The entries in each index correspond to the probability that the match is at that state at that time. Therefore, $\mathbf{x} \in R^{1*21}$, and $\sum_{i=1}^{21} x_i = 1$. Since the match always starts at 'all-square', our initial probability vector, denoted $\mathbf{x}_0$, will have an entry of 1 at state 0, and zeros for all other states.

$$
\begin{array}{c c}
 & \begin{array}{ccccccccccccccccccccc} -10 & -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\
\mathbf{x}_0 = & \left[\begin{array}{ccccccccccccccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]
\end{array}
$$

Figure 5: Initial Stochastic Probability Vector

To calculate the stochastic probability vector after n steps, we use

$$\mathbf{x}_n = \mathbf{x}_0 T^n \tag{6}$$

This is convenient because we can evaluate our stochastic probability vector for any $n = 1, 2, ..., 18$. Therefore, it is very easy to calculate the match score probabilities after any given hole. To find the probabilities of what score the match will end with, we simply set $n = 18$. This gives us

$$
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\left[0.1120,\right. & 0.0743, & 0.1031, & 0.0839, & 0.0596, & 0.0357, & 0.0171, & 0.0060, & 0.0013, & \left.0.0001\right] \\
-10 & -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 \\
\left[0.0001,\right. & 0.0007, & 0.0034, & 0.0104, & 0.0233, & 0.4170, & 0.0631, & 0.0832, & 0.0645, & \left.0.1043\right]
\end{array}
$$

Figure 6: End Score Probabilities for Fowler

0

$$[0.1124]$$

Figure 7: Tie Probability for Fowler

Note that if the reader is just interested in Fowler's overall probability of winning (or losing), and not the probability of the individual states, then

$$P(win) = \sum_{i=1}^{10} x_i \tag{7}$$

$$P(lose) = \sum_{i=-10}^{-1} x_i \tag{8}$$

In our example with Fowler vs Johnson, $P(win) = 0.4931$, $P(lose) = 0.3945$, and of course $P(tie) = 0.1124$.

1.4 Other Works

Wade [1] created a Markov Model that predicts the outcome of a golf match play competition as well. In his model, he first calculates his probabilities in relation to par. I instead decided to use the actual number of strokes it takes a golfer to complete a hole. In a golf sense, I think my way is easier for bookkeeping, because the par changes depending on what hole the golfers are playing. This discrepancy does not change the win/lose/tie percentages, but it does change the probabilities of the individual outcomes. I also believe that this makes my table of probabilities easier to read and compare between the two golfers. This also changes the general formula used to calculate the total probabilities of winning, losing, or tying a specific hole. Another change in my model was to construction of the transition matrix. Wade let his transition matrix only have two absorbing states corresponding to 10-up and 10-down for all 18 holes. However, as I stated in the description of my transition matrix, I added two new absorbing states after each hole starting at hole 11. This way, once a match is finished, it locks the end score of the match and accounts for the probability that the match will end at that state. This way the data isn't slightly skewed by letting the match jump to different scores after it was declared over. Also, I differed by using an initial stochastic probability vector instead of a matrix. This way, every index of the vector corresponds to a score the match could be at. This makes it very easy to see what the individual probability values are for each score of the match. Also, you are not limited to calculating the win/lose/tie probabilities only after 18 holes. With my procedure, you can evaluate the results after any hole.

2 Stroke-Play Golf

2.1 Rules and Scoring

The main difference in stoke-play golf from match-play is that a player's score is simply the total number of strokes (times a player hits the ball) for all 18 holes. This means that in stroke-play competition, there is no limit as to how many competitors there are. If a tournament has 100 golfers competing, then the golfer with the fewest amount of total strokes will place in first. Conversely, the golfer with the most strokes will place in last.

To make scoring comparison more simple, we need to introduce what 'par' for a course is. For every hole, a certain number of strokes is assigned to be a score of 'par' on that specific hole. For example, if hole number 1 is assigned a par of 4, then a golfer that takes 4 strokes to complete the hole is considered to receive a score of par on hole 1. A score of par will correspond to the numerical value 0. If a different golfer needed 5 strokes to complete the hole, then the score assigned is +1. This is because it took this golfer one extra stroke in relation to par to complete the hole. Similarly, a score of -1 will be assigned if the golfer completed the hole with one less stroke in relation to par. The total score is then determined by summing all 18 scores in relation to par. This scoring system is advantageous because at any given hole, it is much easier to evaluate a golfer's performance. Scoring this way does not change the leader board at the end of 18 holes, because score to par is directly proportional to total number of strokes.

2.2 Hole Probabilities

Just like with match-play, I will introduce a simplified version of predicting the outcome of a stroke-play match. I will then expand on this simple model to make a more accurate model.

Unlike match-play, there are more than just three possible outcomes for golfer A for a given hole. Since the lowest score in relation to par possible is -4, and pro golfers almost never score over +4 on a hole, we have 9 possible outcomes. As for notation, A_i will denote that golfer A scored i in relation to par on that particular hole. Since we know that the sum of the probabilities of these events has to equal 1, we have:

$$P(A_{-4}) + P(A_{-3}) + ... + P(A_3) + P(A_4) = 1 \tag{9}$$

For the simple model, it is very easy to calculate these individual probabilities. The $P(A_i)$ will be the proportion of times golfer A scores an i on every previous hole being evaluated. We can look at Rickie Fowler as an example again.

	-4	-3	-2	-1	0	1	2	3	4
Fowler	0	0	0.032	0.2556	0.5228	0.1667	0.021	0.0019	0

Figure 8: Fowler 2017 Hole Score to Par Statistics

2.3 Simple Markov Model

The first step in creating our Markov Model for stroke-play is to construct our right transition matrix T. Since it is very rare for pro golfers to finish 18 holes at score less than -10, or greater than +10, this model will have state space $S = \{-10, -9, ..., 0, ..., 0, 10\}$ and $|S| = 21$. It follows that $T \in R^{21*21}$, because we have 21 states. We then place equation (9) in the correct indices to complete our transition matrix. Since the golfer starts the competition at level par, the stochastic probability vector $\mathbf{x}$ will have value 1 at index 0.

$$\begin{array}{cc} & \begin{array}{ccccccc} -10 & -9 & -8 & & 8 & 9 & 10 \end{array} \\ \begin{array}{c} -10 \\ -9 \\ -8 \\ T = \\ \\ 8 \\ 9 \\ 10 \end{array} & \begin{bmatrix} P(A_0) & P(A_1) & P(A_2) & --- & --- & --- & --- \\ P(A_{-1}) & P(A_0) & P(A_1) & --- & --- & --- & --- \\ P(A_{-2}) & P(A_{-1}) & P(A_0) & --- & --- & --- & --- \\ --- & --- & --- & --- & --- & --- & --- \\ --- & --- & --- & --- & --- & --- & --- \\ --- & --- & --- & --- & P(A_0) & P(A_1) & P(A_2) \\ --- & --- & --- & --- & P(A_{-1}) & P(A_0) & P(A_1) \\ --- & --- & --- & --- & P(A_{-2}) & P(A_{-1}) & P(A_0) \end{bmatrix} \end{array}$$

Figure 9: Transition Matrix

Like before, to calculate the stochastic probability vector after n steps, we use equation (6). Therefore, we can evaluate our probability vector for any hole. It is important to note that although the states are denoted with the same notation as the match play model, they are not the same states. For example, '-4' in

$$
\begin{array}{c}
\begin{array}{ccccccccccccccccccccc} -10 & -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\
\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

Figure 10: Initial Stochastic Probability Vector

the match play model refers to a state of '4-down', while in this model it refers to the state the golfer is at state '-4' in relation to par.

3 Par 3 Forecasting

3.1 Intuition

A golf course will always have some par 3 holes to make up the course. Par 3 holes are the shortest in distance when compared to par 4 and par 5 holes. The idea behind a par 3 hole is for it to be close enough to have a shot at hitting the putting green on the golfer's first shot. My goal is to construct a Markov Chain that shows the step probabilities of how a golfer plays a par 3 hole, which will showcase the golfer's path to the hole and the score probabilities.

3.2 States

We first need to define what the states of the model will be. So after the first shot on a par 3, the ball can either go: on the green, left of the green, right of the green, over the green, short of the green, sand bunker, or water hazard. To be more accurate, we can split up the state of being on the green as: in the hole, on green 0-3 feet from hole, 3-10 feet, 10-30 feet, and 30+ feet. For the golfer's second shot, we can assume all of the same states, but the probabilities will obviously change because it is much easier to hit it on the green from just left of the green as it is from the starting point. The difficult part is determining these probabilities due to the fact all par 3s are slightly different from each other. The biggest differences are length and overall shape of the hole. For a simple example, I will take a par 3 that is 150 yards and straight with only one sand bunker.

	G0	G3	G10	G30	L	R	O	S	B	W
Probability	0	0.08	0.14	0.28	0.11	0.19	0.09	0.06	0.05	0

Figure 11: 150 yards Par 3 First Shot Probabilities

	T	G0	G3	G10	G30	L	R	O	S	B	W
T	0	0	0.08	0.14	0.28	0.11	0.19	0.09	0.06	0.05	0
G0	0	1	0	0	0	0	0	0	0	0	0
G3	0	0.97	0.03	0	0	0	0	0	0	0	0
G10	0	0.61	0.38	0.01	0	0	0	0	0	0	0
G30	0	0.18	0.71	0.1	0.01	0	0	0	0	0	0
L	0	0.02	0.15	0.53	0.38	0.01	0.01	0	0	0	0
R	0	0.04	0.13	0.58	0.32	0.02	0	0	0	0	0
O	0	0.02	0.13	0.6	0.32	0	0	0	0.02	0	0
S	0	0.05	0.2	0.55	0.2	0	0	0	0	0	0
B	0	0.06	0.22	0.51	0.2	0	0	0	0	0.01	0
W	0	0	0	0.05	0.15	0.12	0.18	0.2	0.25	0.01	0.04

Figure 12: 150 yards Par 3 Transition Matrix

T	G0	G3	G10	G30	L	R	O	S	B	W
1	0	0	0	0	0	0	0	0	0	0

Figure 13: 150 yards Par 3 Initial Stochastic Probability Vector

Note that we just have one absorbing state at **G0** , because once the golfer make it in the hole, he does not hit anymore shots. It also makes since to set up the initial vector with just one value at T, because you always start at the tee when hitting your first shot.

3.3 Calculations

Just like before, to find the probability that the golfer lies in each state at step in can be found using:

$$\mathbf{x}_n = \mathbf{x}_0 T^n \tag{10}$$

So to find the values after two shots, we assign $n = 2$. With the current data, this yields:

T	G0	G3	G10	G30	L	R	O	S	B	W
0	0.231	0.3303	0.3104	0.1562	0.0049	0.0011	0	0.0018	0.0005	0

Figure 14: State Probabilities After 2 Shots on Par 3

Therefore, since making it in the hole in two shots counts as a score of -1 (a birdie) we can see we have a 0.231 chance of this happening. This also nicely shows what the probabilities are that we are in the other states after n shots. In more general terms, $\mathbf{G0}_n$ shows the probability that the golfer finished the hole in $\leq n$ strokes. Because of this, it is important to note that the probability that we get par will require more calculation than just looking at the value for $\mathbf{G0}_3$. We will have to subtract the $\mathbf{G0}_3$ probability by the $\mathbf{G0}_2$ probability. This is due to the fact that we want to find the difference in the probability when we move to the next step. This difference denotes the probability we made the ball in the hole on that stroke. We can generalize this approach by producing a formula to calculate the probability value of making it in the hole in exactly n strokes (equation 11). Note that this probability converges to 1 as $n \to \infty$ because the golfer will eventually get the ball in the hole given an infinite amount of shots.

$$P(n) = \mathbf{G0}_n - \sum_{i=1}^{n-1} \mathbf{G0}_i \tag{11}$$

4 Par 4 and 5 Forecasting

4.1 Intuition

The idea behind the Markov model for par 4 scoring is similar to the idea behind the par 3 model I have just described. However, par 4s are generally set up in a way that the golfer is required to hit a tee shot for placement before the golfer can hit at the green like a par 3. This makes the creation of the states for the Markov model a little more difficult to create. I will keep the states from the par 3 model, and just add the additional states needed for a par 4 hole. The pattern follows for par 5 holes as expected. The golfer is required to hit a tee shot to a place that is similar to the start of a par 4 hole. Therefore, for the par 5 states, I will keep the same states as the par 3 and 4 models, with the addition of the states specific to par 5 golf holes.

4.2 Par 4 States

All of the states from the par 3 model will be used as states in the par 4 model. However, I will need to add additional states that account for the tee shot placement before the golfer shoots for the green. So for after the tee shot, I will add states that represent a player being in the fairway (F), left of fairway (FL), right of fairway (FR), fairway sand bunker (FB), and fairway water hazard (FW). The next step will be to add the statistics for the golfer given being at those states. For simplicity, I will assume that the par 4 is set up to where the golfer will have 150 yards to the green after the tee shot. This is shown below, along with the state probabilities after $n = 3$ shots.

	T	G0	G3	G10	G30	L	R	O	S	B	W	FB	FW	FL	FR	F
T	0	0	0	0	0	0	0	0	0	0	0	0.1	0.05	0.22	0.21	0.42
G0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G3	0	0.97	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0
G10	0	0.61	0.38	0.01	0	0	0	0	0	0	0	0	0	0	0	0
G30	0	0.18	0.71	0.1	0.01	0	0	0	0	0	0	0	0	0	0	0
L	0	0.02	0.15	0.53	0.38	0.01	0.01	0	0	0	0	0	0	0	0	0
R	0	0.04	0.13	0.58	0.32	0.02	0	0	0	0	0	0	0	0	0	0
O	0	0.02	0.13	0.6	0.32	0.02	0	0	0	0	0	0	0	0	0	0
S	0	0.05	0.2	0.55	0.2	0	0	0	0	0	0	0	0	0	0	0
B	0	0.06	0.22	0.51	0.2	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0.05	0.15	0.12	0.18	0.2	0.25	0.01	0.04	0	0	0	0	0
FB	0	0	0.03	0.12	0.3	0.18	0.1	0.02	0.2	0.02	0.03	0	0	0	0	0
FW	0	0	0	0	0.02	0.05	0.03	0.01	0.1	0.02	0.01	0	0.12	0.2	0.19	0.25
FL	0	0	0.09	0.19	0.33	0.1	0.11	0.09	0.07	0.01	0.01	0	0	0	0	0
FR	0	0	0.1	0.15	0.35	0.09	0.1	0.04	0.03	0.05	0.09	0	0	0	0	0
F	0	0	0.11	0.2	0.65	0.01	0.01	0.01	0.01	0	0	0	0	0	0	0

Figure 15: Par 4 Transition Matrix

T	G0	G3	G10	G30	L	R	O	S	B	W	FB	FW	FL	FR	F
0	0.279	0.425	0.18	0.091	0.007	0.007	0.006	0.007	0.001	0.002	0	0.001	0.001	0.001	0.001

Figure 16: State Probabilities for $n = 3$

4.3 Par 5 States

Like mentioned before, I will keep the states from the par 4 model. I will add additional states that account for the golfer's tee shot on the par 5. This means a golfer will typically hit a tee shot to the fairway, then 'lay-up' of the green to the fairway again before shooting at the green. Therefore, I will need to differentiate between which part of the fairway the golfer is at so there is no confusion in the model. The part of the fairway the golfer hits to first off the tee will be denoted by '1'. The second part of the fairway the golfer will hit to from part 1 will be denoted '2'. Below is the transition matrix and the state probabilities for $n = 4$.

	T	G0	G3	G10	G30	L	R	O	S	B	W	FB	FW	FL	FR	F	FB1	FW1	FL1	FR1	F1
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0.05	0.22	0.21	0.42
G0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G3	0	0.97	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G10	0	0.61	0.38	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G30	0	0.18	0.71	0.1	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0.02	0.15	0.53	0.38	0.01	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0.04	0.13	0.58	0.32	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0.02	0.13	0.6	0.32	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0.05	0.2	0.55	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0.06	0.22	0.51	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0.05	0.15	0.12	0.18	0.2	0.25	0.01	0.04	0	0	0	0	0	0	0	0	0	0
FB2	0	0	0.03	0.12	0.3	0.18	0.1	0.02	0.2	0.02	0.03	0	0	0	0	0	0	0	0	0	0
FW2	0	0	0	0	0.02	0.05	0.03	0.01	0.1	0.02	0.01	0	0.12	0.2	0.19	0.25	0	0	0	0	0
FL2	0	0	0.09	0.19	0.33	0.1	0.11	0.09	0.07	0.01	0.01	0	0	0	0	0	0	0	0	0	0
FR2	0	0	0.1	0.15	0.35	0.09	0.1	0.04	0.03	0.05	0.09	0	0	0	0	0	0	0	0	0	0
F2	0	0	0.11	0.2	0.65	0.01	0.01	0.01	0.01	0	0	0	0	0	0	0	0	0	0	0	0
FB1	0	0	0	0	0	0	0	0	0	0	0	0.18	0.09	0.2	0.1	0.33	0.05	0.05	0	0	0
FW1	0	0	0	0	0	0	0	0	0	0	0	0.07	0.01	0.09	0.08	0.12	0.1	0.08	0.12	0.11	0.22
FL1	0	0	0	0	0	0	0	0	0	0	0	0.09	0.08	0.19	0.11	0.53	0	0	0	0	0
FR1	0	0	0	0	0	0	0	0	0	0	0	0.08	0.06	0.22	0.17	0.47	0	0	0	0	0
F1	0	0	0	0	0	0	0	0	0	0	0	0.06	0.04	0.11	0.12	0.67	0	0	0	0	0

Figure 17: Par 4 Transition Matrix

T	G0	G3	G10	G30	L	R	O	S	B	W	FB2	FW2	FL2	FR2	F2	FB1	FW1	FL1	FR1	F1
0	0.277	0.434	0.159	0.093	0.008	0.007	0.005	0.007	0.001	0.002	0.001	0.001	0.002	0.002	0.005	0	0	0	0	0

Figure 18: State Probabilities for $n = 4$

Note that we can refer back to equation (11) for the par 4 and par 5 models as well.

5 Conclusion

I have developed and discussed five different golf related Markov models in this paper. The first model calculated the probabilities at any hole for a match-play golf competition between two golfers. The second model focused on forecasting a single golfer's score at any hole in stroke-play competition. The last 3 models all pertained to forecasting how a specific golfer will play any singular hole depending on what par the hole is. These models are able to give the probabilities of where the golfer's ball could be after any amount of shots.

Bibliography

[1] Shasky, Wade. *Markov Chains and its Application to Golf*, Lakehead University, 2015.
[2] “Golf Stats - PGA TOUR.” PGATour, www.pgatour.com/stats.html.