Net Best-Ball Team Composition in Golf

by

Yifan (Lucas) Wu

B.Sc., Simon Fraser University, 2017

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2018

Approval

Name:	**Yifan (Lucas) Wu**
Degree:	**Master of Science (Statistics)**
Title:	**Net Best-Ball Team Composition in Golf**
Examining Committee:	**Chair:** Dr. Jiguo Cao Associate Professor
	Dr. Tim B. Swartz Senior Supervisor Professor
	Dr. Harsha Perera Supervisor Lecturer
	Dr. Peter Chow-White External Examiner Professor School of Communication
Date Defended:	**August 7, 2018**

Abstract

This project proposes a simple method of forming two-player and four-player golf teams for the purposes of net best-ball tournaments in stroke play format. The proposal is based on the recognition that variability is an important consideration in team composition; highly variable players contribute greatly in a best-ball setting. A theoretical derivation is provided for the proposed team formation. In addition, simulation studies are carried out which compare the proposal against other methods of team formation. In these studies, the proposed team composition leads to competitions that are more fair.

Keywords: Sports analytics, Distribution of minimum, Handicapping in golf, Order statistics, Simulation

"I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I-
I took the one less traveled by,
And that has made all the difference."
The Road Not Taken
by Robert Frost

Acknowledgements

I would first like to express my deepest gratitude to my senior supervisor Dr. Tim Swartz for his continual support, guidance and patience throughout my journey at Simon Fraser University. Without his encouragement, I would not switch my major from Business to Statistics in the third year of my undergraduate study. Thank you for placing trust in me when I was a student with little background in Statistics.

Thank you to all the members of my examining committee, Dr. Jiguo Cao, Dr. Peter Chow-White and Dr. Harsha Perera for their valuable suggestions and feedbacks. I want to thank Peter for introducing me to the field of Sports Analytics and having me in the GeNA Lab.

I am grateful to all the faculty members who taught me during my undergraduate and graduate studies, including Dr. Rachel Altman, Dr. Luke Bornn, Dr. Dave Campbell, Dr. Jinko Graham, Dr. Tom Loughin, Dr. Brad McNeney, Dr. Carl Schwarz, Dr. Tim Swartz and Dr. Boxin Tang. I also wish to thank Charlene Bradbury, Sadika Jungic and Kelly Jay for their kind assistance and help. Special thanks to Robin Insley, Marie Loughin and Ian Berjovitz for their support and help during my time at SFU.

I would also like to thank my fellow students and friends including my sports analytics crew, office mates, foodie mates, etc., for all the fun we had along the way.

Last but not least I would like to thank my family, especially my parents for their unconditional love, support and understanding.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Golf can be played in a variety of formats including stroke play, match play and best-ball. In stroke play, the players compare their total number of strokes taken to sink the ball over 18 holes and the golfer who takes the fewest strokes is the winner. In match play, the competition is played on a hole-by-hole basis. A player wins a hole by finishing that hole in fewer strokes than the other players. The match play winner is the player who wins the most number of holes. In best-ball format, golfers play in teams and the best (lowest) score for a team on a hole is determined by the best (lowest) score of the team members.

In most sports, it would be almost impossible for amateurs to compete with professional players such as LeBron James in basketball and Lionel Messi in soccer. However, golf is one of the few sports that allows players with different skill levels to compete against each other, which means you might have a shot to defeat Tiger Woods in golf. This is achieved by the handicapping system which allows for fair competition between any two players (Yun 2011a, 2011b, 2011c, 2011d).

The fairness of the handicapping system used in Canada and the United States has been studied extensively in the literature. Some work suggests that in a head-to-head competition between two golfers, the lower handicap player (the stronger one) is slightly favored in both stroke and match play formats (Madras 2017, Kupper et al. 2012, Bingham and Swartz 2000, Scheid 1977 and Pollock 1974). In Section 10-2 of the United States Golf Association (USGA) Handicap System Manual (USGA 2016), it is stated that the handicap formula provides an “incentive for players to improve their golf games” whereby a small bonus for excellence is given to the stronger player.

However, solely relying on the current handicapping system cannot guarantee fairness when multiple golfers play in team formats with different rules. For instance, Bingham and Swartz (2000) suggested that weaker golfers have a considerable advantage in winning a tournament based on net scores. Grasman and Thomas (2013) investigated scramble competitions and proposed a method which aims to provide all scramble teams with equal expectations of winning the tournament. It was suggested by Hurley and Sauerbrei (2015) that team net best-ball matches are generally unfair when all golfers are assigned their full handicap and they stated that they had been unable to find a simple rule to make a net best-ball competition fair. Hurley and Sauerbrei's (2015) research shows that team net best-ball matches are not generally fair, which motivates this paper to find new ways to improve the fairness in net best-ball competition.

There are varying handicapping systems in different countries. The United States Golf Association (USGA) and Golf Canada have made efforts to develop handicapping systems that make the game fair (USGA 2016, Golf Canada, 2016). McHale (2010) also discussed alternative handicapping systems used in the United Kingdom and other European countries. In this paper, we focus on the handicapping system used by the USGA and the Royal Canadian Golf Association (RCGA). It is also noteworthy that fairness is defined differently in the literature. For example, Benincasa, Pavlikov, Hearn (2017) proposed two different definitions of fairness. One definition requires equal probabilities of each team to win the match. For the purpose of this paper, we define fairness as equal probabilities of all teams finishing in a fixed position. In another words, a fair system involving n teams is one where the probability of each team finishing in kth place is $1/n$, $k = 1, \dots, n$.

There are two relevant papers that concern the problem of team formation in net best-ball competitions. Siegbahn and Hearn (2010) studied a two-player versus two-player event that involves handicapping. Similar to what was presented in Bingham and Swartz's paper in 2000, golfer variability was a main focus of their study such that the variability of golfer performance was evaluated and estimated through a function of handicap. Siegbahn and Hearn (2010) concluded that high handicap golfers (i.e. weak golfers) have an advantage in fourball and they proposed some tie-breaking rules to make the game more fair. As discussed in Siegbahn and Hearn (2010), previous studies on fairness in fourball matches focused on the difference in handicaps between teammates as a predictor of fourball success.

Pavlikov, Hearn and Uryasev (2014) carried out their research based on the results of Siegbahn and Hearn (2010) to address the issue of team composition in net best-ball tournament settings. They proposed a search algorithm over the combinatorial space of potential team compositions. Optimal team formations were determined based on the assumption that all teams have approximately equal probability of winning. Their search algorithm can be executed most efficiently when the number of golfers $n < 40$. One advantage of using this algorithm is that it can be run on any prescribed team size. However, a major disadvantage of this approach is that it relies on the use of tables that contain average scoring distributions for players of a given handicap. This implies an monotonic relationship between handicap and variability. High handicap golfers tend to have a higher variability in general. In reality, there are instances of high handicap golfers who are consistent (ie. less variable). For example, there exists senior golfers may not hit the ball far, but are straight off the tee and rarely get into trouble (i.e. the rough, hazards, water, etc.). Additionally, the dataset used in Chapter 4 suggests that there is not a strictly monotonic relationship between handicap and variability. Therefore, instead of using the average characteristics of golfers, our approach estimates variability individually for golfers following Swartz (2009). This forms the critical component for our team formation proposal in net best-ball tournaments.

In Chapter 2, we review background material that is related to team composition. This includes details concerning the rules related to net best-ball tournaments, the current handicapping system and the related literature. In Chapter 3, we discuss our proposed team composition for net best-ball competitions. The key idea is based on the recognition that variability is a crucial determinant in terms of team performance in net best-ball competitions. The method is to pair golfers of high variability with golfers of low variability, which can be used for both two-man and four-man team competitions. More notably, the procedure is simple to implement and does not require any special software. In Chapter 4, two simulation studies are provided. The first study uses a theoretical model for golf scores to examine the performance of all possible team compositions. The second study investigates the common practices of team composition using a resampling procedure of actual golf scores. In the results from both studies, our team composition method contributes to fairer competitions. A short discussion is provided in Chapter 5 to conclude our study. The material in this MSc project is an expansion of Wu, Chow-White and Swartz (2018).

Chapter 2

Background Material

We proposed a simple and easy method for net best-ball team composition. Before we dive into the details of our proposed method, we need to introduce some background material, which provides the rationale and structure for the proposed method.

2.1 Net best-ball competitions

Net best-ball competitions are typically based on teams of size $m = 2$ or teams of size $m = 4$. And in such competitions, we denote that there are $n \geq 2$ teams. Golf can be played on a 9-hole or 18-hole course. The majority of competitions are carried out on an 18-hole course.

Consider a competition with 18-holes. On the jth hole of the course, $j = 1, \ldots, 18$, the ith player on a given team has a *gross score* X_{ij} which represents the number of strokes that it took him to hit the ball into the hole. Associated with the ith player on the jth hole is a handicap allowance $h_{ij} = 0, 1, 2$ which is related to the quality of the player. The larger the value of h_{ij}, the weaker the player. The weaker player should get more strokes deducted when competing against a stronger player. Also, handicap allowance varies with different forms of competition and attempts to produce equitable competitions (USGA 2016). Under this framework, the ith player has the resultant *net score*

$$Y_{ij} = X_{ij} - h_{ij} \tag{2.1}$$

on hole j. The player's team then records their *net best-ball score* on the jth hole

$$T_j = \min_{i=1,\ldots,m} Y_{ij} \tag{2.2}$$

and the team's overall performance is based on their aggregated net best-ball score

$$T = \sum_{j=1}^{18} T_j \ . \tag{2.3}$$

The teams in the competition are then ranked according to (2.3) where the winning team has the best (lowest) value of T. If teams have the same T, there exist various ways for breaking ties. For example, with multiple ties, the team having done best on the 18th hole may be determined the winner. If a tie still exists, the criteria may then be based on the 17th hole, then the 16th hole, etc.

The above format is known as *stroke play* which is the focus of our investigation. When there are only $n = 2$ teams, then *match play* competitions are possible. In match play, T_j is calculated as in (2.2), and the team with the lower value of T_j is said to have won the jth hole. At the end, the team with the greatest number of winning holes is the match play winner.

2.2 The current handicapping system

Details of how to calculate handicap under different scenario are described in section 10 of the USGA Handicap System Manual (USGA 2016). Handicap system is a complicated system, so we attempt to provide a description of the standard calculation which applies to most golfers.

Consider then a golfer's most recent 20 rounds of golf where each round is completed on a full 18-hole golf courses. The kth round yields the *differential* D_k which is obtained by

$$D_k = \frac{(\text{adjusted gross score} - \text{course rating}) * 113}{\text{slope rating}} \tag{2.4}$$

In (2.4), the *adjusted gross score* is the player's actual score reduced according to *equitable stroke control* (ESC) which is a mechanism for limiting high scores on individual holes. The intuition is that handicap reflects potential and should not be distorted by unusually poor results. The *course rating* describes the difficulty of the course from the perspective of a *scratch* (expert) golfer. Typically, course ratings are close to the *par* score of the course where values less than (greater than) par indicates less (more) difficult courses. Course ratings are reported to one decimal place. The *slope rating* describes the difficulty of the course from the perspective of non-scratch golfers where a slope rating less than (greater than) 113 indicates an easier (more difficult) course than average. Slope ratings are integer-valued and lie in the interval $(55, 155)$. The main takeaway from (2.4) is that large differentials correspond to poor rounds of golf and small differentials correspond to excellent rounds of golf. It is even possible for differentials to be negative. Differentials are rounded to the first decimal place.

Given a golfer's scoring record, the golfer's *handicap index* is calculated by taking 96% of the average of the 10 best (lowest) differentials and truncating the result to the first decimal place. In Section 2.3, we will see that it is instructive to write the handicap index as

$$\begin{aligned} I &= (0.96)(D_{(1)} + D_{(2)} + \cdots + D_{(10)})/10 \\ &= (0.096)D_{(1)} + (0.096)D_{(2)} + \cdots + (0.096)D_{(10)} \end{aligned} \tag{2.5}$$

where $D_{(i)}$ denotes the ith order statistic of the differentials. The handicap index (2.5) is the summary statistic that is used in USGA handicapping; small values indicate strong golfers whereas large values indicate weak golfers. It is possible that the handicap index $I < 0$, and these golfers (mostly professionals) are referred to as *plus* golfers. The maximum allowable handicap index for men is 36.4. For many golfers, the calculation of the handicap index I is viewed as a black-box procedure. Under the RCGA jurisdiction (Golf Canada 2016), handicap index is referred to as *handicap factor*.

Recognizing that courses are of varying difficulty, the last step for the implementation of handicap involves converting the handicap index to strokes for a particular course. For a course with slope rating S, the *course handicap* for a golfer with handicap index I is given by $C = I * S/113$ rounded to the nearest integer.

In the context of net best-ball competitions, the course handicaps C of the players in the tournament are then used to determine the hole-by-hole handicap allowances h_{ij} in (2.1). It is at this point where there is some variation in how h_{ij} is obtained. According to Section 9-4(bii) of the USGA Handicap System Manual (USGA 2016), the recommended way is to first reduce the individual course handicaps C by a factor of 90%, rounding to the nearest integer. An adjustment is then made to the reduced course handicaps where the course handicap for a given golfer is set to the offset between their course handicap and the lowest (best) course handicap in the competition. For example, suppose that the best golfer in the competition has reduced course handicap $C_{i_1} = 3$ and that some other golfer has reduced course handicap $C_{i_2} = 24$. Then the two course handicaps are converted to $C_{i_1} = 3 - 3 = 0$ and $C_{i_2} = 24 - 3 = 21$, respectively. Then, we note that the holes on a golf course are assigned a *hole handicap* according to a stroke allocation table. The table consists of a permutation of the integers 1 to 18 where it is typically thought that increasing numbers correspond to decreasing difficulty of the holes. Denote the hole handicap on the jth hole by HDCP_j. Under the complicated framework described above, h_{ij} is determined as follows:

$$h_{ij} = \begin{cases} 0 & \text{HDCP}_j > C_i, & C_i \le 18 \\ 1 & \text{HDCP}_j \le C_i, & C_i \le 18 \\ 1 & \text{HDCP}_j > C_i - 18, & C_i > 18 \\ 2 & \text{HDCP}_j \le C_i - 18, & C_i > 18 \end{cases}. \tag{2.6}$$

Although (2.6) may be difficult to digest, the idea is that relative to the strongest player, an individual with $C \le 18$ receives a single stroke on the most difficult holes up to his handicap offset. If his handicap offset exceeds 18, then he receives two strokes on the more difficult holes and one stroke on the remaining holes.

Madras (2017) investigated how alternative permutations of HDCP and other innovations affect the fairness of net match play events between two players.

2.3 Related literature and ideas

In consultation with the RCGA, Swartz (2009) proposed an alternative handicapping system with the following features:

- the system retains the well-established concepts of course rating and slope rating
- the system provides a modified handicap index/factor referred to as the *mean* which has a clear interpretation in terms of actual golf performance; this is contrasted with the index/factor whose interpretation is related to potential
- the system was developed using probability theory, leading to net competitions that are more fair

The key component of the system developed by Swartz (2009) was that it incorporated variability in handicapping. And in the context of net best-ball tournaments, it is clear that amongst two golfers with the same handicap index, a highly variable golfer is more valuable to a team than a consistent golfer. For example, the highly variable golfer will obtain more net birdies which contribute positively to the overall net score of his team. On the other hand, when this highly variable golfer scores net double bogeys, these poor scores are not likely to penalize his team in the best-ball format.

As an alternative to the handicap index/factor, Swartz (2009) defined two statistics that characterize player performance. These statistics are referred to as the mean $\hat{\mu}$ and the spread $\hat{\sigma}$, and their calculation is analogous to (2.5). Specifically,

$$\hat{\mu} = w_1 D_{(1)} + w_2 D_{(2)} + \cdots + w_{16} D_{(16)} \tag{2.7}$$

and

$$\hat{\sigma} = q_1 D_{(1)} + q_2 D_{(2)} + \cdots + q_{16} D_{(16)} \tag{2.8}$$

where the weights w_i in (2.7) and q_i in (2.8) provide best linear unbiased estimators (BLUEs) of the mean and the standard deviations of the differentials where the differentials are assumed to be realizations of independent and identically distributed normal random variables. Whereas (2.5) is based on 10 order statistics, (2.7) and (2.8) are based on 16 order statistics; the rationale was that data is informative and it is wasteful to discard observations. On the other hand, there is evidence that the largest differentials may not arise from a normal distribution as the true underlying distribution may be positively skewed (Siegbahn and Hearn 2010).

For the purposes of this paper, the spread $\hat{\sigma}$ in (2.8) plays a primary role and we record the weights q_i in Table 2.1. Alternative weights are recorded in Swartz (2009) when a golfer has played fewer than 20 complete rounds. When the spread $\hat{\sigma}$ falls outside of the interval (1.5,8.0), it is set equal to the corresponding endpoint.

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
-0.1511	-0.1006	-0.0792	-0.0632	-0.0500	-0.0384	-0.0277	-0.0178
q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
-0.0082	0.0011	0.0103	0.0196	0.0291	0.0389	0.0492	0.3880

Table 2.1: The weights q_i in (2.8) that are used in the calculation of the spread $\hat{\sigma}$.

A point that is worth emphasizing is that the calculation of $\hat{\sigma}$ in (2.8) is simple and is directly analogous to the calculation of (2.5) which is part of the current handicapping system. In Section 3, we assume that the values $\hat{\sigma}$ are available for each golfer in the net best-ball competition. Moreover, the values $\hat{\sigma}$ are the only values that are needed to form teams according to our proposal.

Chapter 3

Team Composition

Consider a net best-ball tournament where the number of competitors is even. Initially, our task is to pair players in a fair manner. Following (2.1), we let Y_{ij} denote the net score of golfer i on hole j. Although golf scores are discrete, we assume

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \tau_{ij}^2) \ . \tag{3.1}$$

There is some literature that supports the approximate normality of aggregate golf scores (Scheid 1990; Bindham & Swartz, 2000). Scheid (1990) found that golf scores for a round are approximately normally distributed except for a slightly longer right tail. The purpose of handicapping is to create equitable matches. Therefore, we make the assumption that all golfers have the same mean net scores after applying the handicapping adjustment, i.e. $\mu_{ij} = \mu_j$. This implies that (3.1) simplifies to

$$Y_{ij} \sim \text{Normal}(\mu_j, \tau_{ij}^2) \ . \tag{3.2}$$

In addition, we are going to further assume that the mean net scores and the net score variances are the same on all holes, which leads to $\mu_j = \mu$ and $\tau_{ij} = \tau_i$. This is not a very realistic assumption. However, the same analysis can be undertaken on a hole-by-hole basis leading to the same proposal for team compositions. This implies that (3.2) simplifies to

$$Y_{ij} \sim \text{Normal}(\mu, \tau_i^2) \ . \tag{3.3}$$

Without loss of generality, we also set $\mu = 0$ as it is only comparative golf scores that are relevant. Accordingly, we simplify (3.3) whereby the net score for golfer i on each hole is given by

$$Y_i \sim \text{Normal}(0, \tau_i^2) \ . \tag{3.4}$$

With a two-man team consisting of players i_1 and i_2, the quantity of interest is the distribution of the net best-ball result

$$Z_{i_1,i_2} = \min(Y_{i_1}, Y_{i_2}) \ .$$

It is shown by Nadarajah and Kotz (2008) that Z_{i_1,i_2} is nearly normal if τ_{i_1} and τ_{i_2} are close. Using (3.4) and assuming independence between Y_{i1} and Y_{i2}, the moment expressions (11) and (12) from Nadarajah and Kotz (2008) lead to the approximate distribution

$$Z_{i_1,i_2} \sim \text{Normal}\left(-\frac{1}{\sqrt{2\pi}}(\tau_{i_1}^2 + \tau_{i_2}^2)^{1/2} \ , \ \left(\frac{\pi - 1}{2\pi}\right)(\tau_{i_1}^2 + \tau_{i_2}^2) \right) \ . \tag{3.5}$$

If we pair golfers such that every pair has the same probability distribution, then each pair has the same probability of finishing in any position in a tournament. Therefore, if the i_1's and i_2's are paired such that $\tau_{i_1}^2 + \tau_{i_2}^2 = c^2$ for some constant c, then the objective is achieved as each distribution in (3.5) is Normal$(-c/\sqrt{2\pi}, c^2(\pi - 1)/(2\pi))$.

Therefore, we have a prescription for pairing golfers in two-man net best-ball tournaments. We use $\hat{\sigma}$ in (2.8) as a proxy for τ, and we simply match the golfer with the highest $\hat{\sigma}$ with the golfer with the lowest $\hat{\sigma}$, we match the golfer with the second highest $\hat{\sigma}$ with the golfer with the second lowest $\hat{\sigma}$, and so on. Given the $\hat{\sigma}$ values, the forming of two-man teams is an incredibly easy task for the golf director. The matching procedure provides means in (3.5) that are as similar as possible.

In the case of four-man net best-ball tournaments, our heuristic is to begin with optimal two-man teams as described above, and then combine pairs of the two-man teams. It therefore makes sense to form teams based on the mean values in (3.5). Our procedure is to rank the two-man teams according to $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$. We then match the two-man team with the highest $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$ with the two-man team with the lowest $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$, and so on. Again, given the $\hat{\sigma}$ values, this is an incredibly easy task for the golf director.

We have therefore provided a theoretical justification for team composition. In Chapter 4, we supplement the theoretical derivation by simulation studies.

Chapter 4

Simulation Studies

A theoretical justification of our proposed team composition was demonstrated in Chapter 3. We would also like to investigate the performance of our method under different simulation settings. We first generate golf scores from a theoretical model for scoring. We then use a resampling method to generate golf scores from a real data set.

4.1 Simulation via a theoretical scoring model

The first simulation is based on a theoretical model for golfers with different net score probabilities. In Table 4.1, we provide a net score probability distribution of getting a birdie(-1), par(0), bogey(1), double-bogey(2) for 8 golfers. For instance, the eighth row in the table can be interpreted as the chance of getting a birdie, par, bogey, double-bogey for Golfer #8. The percentages are $32\%, 41\%, 22\%$ and 5%, respectively. These 8 golfers are constructed by having the same mean of zero but with an increasing variability. As seen from the table, golfer 1 is the most consistent and golfer 8 is the most variable. We do not believe that these probability distribution are realistic; the intention is to study golfers who have different degrees of variability. The common mean is sensible as this is what is expected of net scores via a fair handicapping system.

Golfer	Net Score Probability Distribution				Mean	SD
	-1	0	1	2		
1	0.02	0.96	0.02	0	0	0.2000
2	0.06	0.88	0.06	0	0	0.3464
3	0.1	0.8	0.1	0	0	0.4472
4	0.14	0.72	0.14	0	0	0.5292
5	0.18	0.64	0.18	0	0	0.6000
6	0.2	0.61	0.18	0.01	0	0.6481
7	0.26	0.51	0.2	0.03	0	0.7616
8	0.32	0.41	0.22	0.05	0	0.8602

Table 4.1: Net score probability distributions of 8 fictitious golfers arranged in an ascending order of standard deviation (SD).

Imagine these 8 golfers are competing in teams of size 2 whose net score abilities follow the probability distribution in Table 4.1. With 8 golfers, we form 4 teams (each with 2 golfers). In this case, the number of possible combinations of team formations is equal to $\frac{\binom{8}{2}\times\binom{6}{2}\times\binom{4}{2}\times\binom{2}{2}}{4!} = 105$ combinations. For each tournament (i.e. simulation), we first generate 18 hole scores for each golfer using the table. We then determine each team's best net score for all 18 holes. Once we have the best net score for each hole, we record the cumulative net score over 18 holes for each team. Finally, we determine which team finishes 1st, 2nd, 3rd or 4th by ranking the cumulative score of each team. We repeat this tournament process 1000 times to obtain frequency tables. Note that if a tie exists in a tournament, we randomly assigned a winner between the ties for that particular tournament. In the long run, this minimizes the bias introduced by different heuristic approaches handling ties.

According to our theory in (3.5), the optimal pairing would match the golfer with the highest $\hat{\sigma}$ with the golfer with the lowest $\hat{\sigma}$ and so on. According to the standard deviation (SD) reported in Table 4.1., this leads to one of the possible team compositions which is 1&8, 2&7, 3&6, 4&5. We denote this theoretical team pairing as WCS (an acronym based on the author's surnames). We are interested in how WCS performs compared to the other 104 pairings in the simulation study?

Recall that our objective is to find a fair system where the probabilities of each team finishing in kth place is 1/4, $j = 1, \ldots, 4$. We could produce a square table of frequencies where rows correspond to finishing order and columns refer to teams. Given 1000 simulations, a fair tournament would give expected cell entries $1000/4 = 250$. Table 4.2 is an example of a 4 by 4 table based on 1000 simulations using the theoretical team pairing, WCS. We observe for example that 1&8 is the strongest team and it finished first 374 times out of the 1000 tournament simulations.

	1&8	2&7	3&6	4&5
Finished 1st	374	247	177	202
Finished 2nd	285	254	206	255
Finished 3rd	215	270	278	237
Finished 4th	126	229	339	306

Table 4.2: Frequencies of the simulated finishing positions corresponding to WCS.

However, whether WCS is meritorious can only be determined in the context of the other 104 potential tournament constructions. And each tournament construction has a corresponding Table 4.2 resulting from the simulation procedure. For each tournament construction, it is natural to assess fairness via the Chi-square test statistic

$$\tilde{\chi}^2 = \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{(f_{ij} - E_{ij})^2}{E_{ij}} \tag{4.1}$$

where $E_{ij} = 250$ and the frequency f_{ij} is the (i, j)th entry from the corresponding matrix as in Table 4.2. Under the null hypothesis that the tournament construction is fair, $\tilde{\chi}^2$ has a Chi-square distribution on 9 degrees of freedom. Large values of $\tilde{\chi}^2$ provide evidence against the null hypothesis.

Table 4.3 provides the Chi-square statistic based on (4.1) for each of the 105 team formation. A list of top 5 and bottom 5 pairings ranked by Chi-square statistic is shown in Table 4.3. The best pairing that gives the lowest Chi-square statistic is the theoretical team formation, WCS. The top 5 methods are all similar in a sense that the golfer with the highest variability is paired with ones with low variability. Conversely, the last row 1&2, 3&4, 5&6, 7&8, leads to significant unequal team variance among 4 teams. And it turns out to perform the worst out of 105 team formations. Therefore, the simulation provided a theoretical justification for team formation by considering variability when constructing teams. We note that the Chi-square statistic is strongly significant for each of the 105 team formations. This indicates that no team formation is "fair". However, our objective is to devise a method that is as fair as possible given the golfers. The simulation exercise demonstrates the WCS is the most fair.

The traditional way of pairing golfers in this context is to match the highest handicap golfer with the lowest handicap golfer, the second highest handicap golfer with the second lowest handicap golfer, and so on. Referring to Table 4.1, we have created golfers with the corresponding net score distributions. It is unclear what the traditional pairing would be in this case because we do not know the gross score distributions (ie. we do not know the strength of the golfers).

Ranking	Team A	Team B	Team C	Team D	Chi-square test statistic
1 (WCS)	1 & 8	2 & 7	3 & 6	4 & 5	221.38
2	1 & 8	2 & 7	3 & 5	4 & 6	270.39
3	1 & 7	2 & 8	3 & 6	4 & 5	371.23
4	1 & 8	2 & 6	3 & 7	4 & 5	420.78
5	1 & 7	2 & 8	3 & 5	4 & 6	432.97
...	...	...	...	...	...
101	1 & 2	3 & 6	4 & 5	7 & 8	3704.98
102	1 & 2	3 & 4	5 & 7	6 & 8	3738.70
103	1 & 2	3 & 4	5 & 8	6 & 7	3851.50
104	1 & 2	3 & 5	4 & 6	7 & 8	3919.31
105	1 & 2	3 & 4	5 & 6	7 & 8	4046.97

Table 4.3: The five best and five worst team formations according to the Chi-square test statistic in (4.1) in an ascending order.

4.2 Simulation via actual golf scores

We now investigate the fairness of team formation using individual handicap factors and hole-by-hole scores. In our analysis, we consider a data set from Coloniale Golf Club in Beaumont, Alberta from 1996 to 1999. This data set is particularly suitable for our simulation since it includes a large pool of golfers with drastically different skill levels. It should therefore have the required generality for testing our proposed method of team formation in net best-ball tournaments.

Prior to our analysis, we pre-process the Coloniale data by retaining golfers who played at least a certain number of rounds. The dataset consists of both male and female golfers. We only keep male golfers in our analysis for two reasons. Firstly, the handicapping calculation is different for male and female golfers. Secondly, the sample size of female golfers who played at least 40 rounds is very small. After restricting the analysis to male golfers who have played at least 40 rounds, we are left with a data set consisting of 10,470 rounds collected on 80 golfers. In Figure 4.1, we provide a histogram of the handicap differential (2.4) corresponding to the 10,470 rounds at Coloniale. The mean handicap differential is approximately 13 and is marked by the dashed red line. We can also observe that golfing ability varies greatly with differentials ranging from -15 to 40.

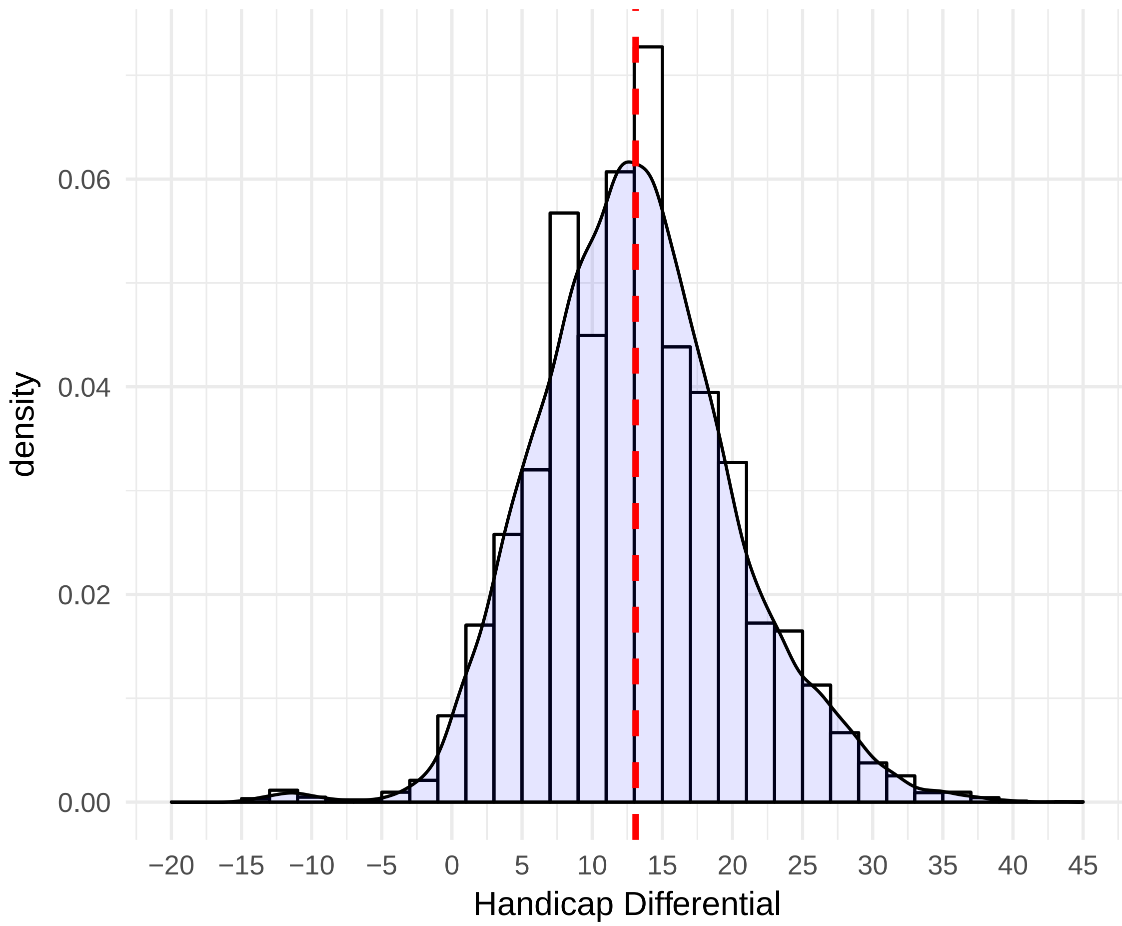

Figure 4.1: Histogram and density plot of the 10,470 handicap differentials corresponding to 80 golfers at the Coloniale Golf Club.

We used a resampling-based method for our simulation procedure. For these 80 golfers, suppose that we investigate a given team construction where we form $n = 20$ teams of size 4. Our simulation can be broken down into two main steps. For each golfer, we first randomly draw 20 of his rounds of golf to calculate his handicap index I in (2.5) and his spread statistic $\hat{\sigma}$ in (2.8). With these two statistics, we are able to compare WCS with other common team formation methods in practice. Secondly, for each of the 80 golfers, we draw one round of hole-by-hole scores from their remaining rounds of golf. It is rare to find detailed hole-by-hole scores, which is a special feature of our dataset. Once we have the generated round of golf for each of the 80 golfers, there are several adjustments made to determine the net score for each golfer in (2.1), such as handicap allowance for golfers playing from different tees. The detailed rules are documented in Section 9-4 of the Golf Canada Handicap Manual (Golf Canada, 2016). After we have determined the net score for each golfer, each team's aggregate net best-ball score can be obtained according to (2.3), and we can further obtain the finishing order of the 20 teams composed by the particular team construction heuristic. This resampling procedure is repeated over 40,000 hypothetical tournaments. Frequency tables are obtained as in Table 4.2 where finishing order corresponds to the rows and team constructions correspond to the columns. In this simulation setting, we have matrices of dimension 20 $\times$ 20.

Recall from Chapter 3 that our method of team formation first ranks golfers according to $\hat{\sigma}$ and then pairs golfers 1&80, 2&79, and so on in a 2-man team. Then, these 40 pairs (i.e. 2-man teams) are ranked according to $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$ where i_1 and i_2 are in the same pair. We then pair the pairs (i.e. 2-man teams) as before with high values of $\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2$ matched with low values. In this way, we are able to determine the 4-man teams.

We now compare our proposed method WCS against two heuristic team formations - "High-Low" and "Zigzag". One of the most common methods of team formation is referred as "High-Low", where High-Low is very similar in construction to WCS. The only difference is that orderings are based on the handicap index I in (2.5) rather than spread statistic $\hat{\sigma}$ in (2.8). The intuition behind the High-Low heuristic is that strong golfers are matched with weak golfers in the first pairing, and then strong teams (based on cumulative handicap indices) are matched with weak teams in the subsequent pairing.

Another method of team formation is referred as "Zigzag", which is viewed as more complex but an improvement over High-Low. Pavlikov, Hearn and Uryasev (2014) provide an illustrative example of Zigzag. For simplicity, consider the formation of 16 golfers into four teams of four players as shown in Table 4.4. Similar to High-Low, golfers are first ordered by handicap index from the lowest to highest. Then match golfers as demonstrated in Table 4.4. For example, Team 1 consists of golfers 1, 8, 9 and 16. The intuition behind Zigzag is similar to High-Low construction but structurally different.

	Golfer ordered by handicap index I from the lowest to highest															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Team 1	x							x	x							x
Team 2		x					x			x					x	
Team 3			x			x					x			x		
Team 4				x	x							x	x			

Table 4.4: Demonstration of the Zigzag method of team composition where 16 golfers are formed into 4 teams.

In Table 4.5, we provide the results of the comparison using the simulated frequency tables based on High-Low, Zigzag and WCS. We calculate the Chi-square statistic (4.1) of these three methods of team formation based on 40,000 simulated tournaments. In this case, the summations in (4.1) extend over the 20 × 20 cells and $E_{ij} = 2{,}000$. We observe that WCS outperforms the other two methods in terms of giving the lowest Chi-square statistic. The High-Low method gives the highest Chi-square statistic, which indicates it is the worst of the three methods in terms of fairness.

	Chi-square
High-Low	4049.09
Zigzag	2136.88
WCS	2034.99

Table 4.5: Comparison of the three methods of team formation based on the Chi-square test statistic.

Besides using the Chi-square statistic, we could also assess the performance of these three methods is via heatmaps. In Figure 4.2, we produce heatmaps corresponding to the simulated frequency tables based on High-Low, Zigzag and WCS. Recall that $E_{ij} = 2{,}000$. Ideally, the fairest competition occurs when each cell has a value of 2000, and the heatmap should have the same color of 2000 across all the cells visually. A deviation from 2000 (i.e. less fair) indicates a more uneven coloring in the heatmaps. In our simulation setting, it is evident that WCS has a more consistent coloring; this indicates that it is a fairer method of team composition than both High-Low and Zigzag.

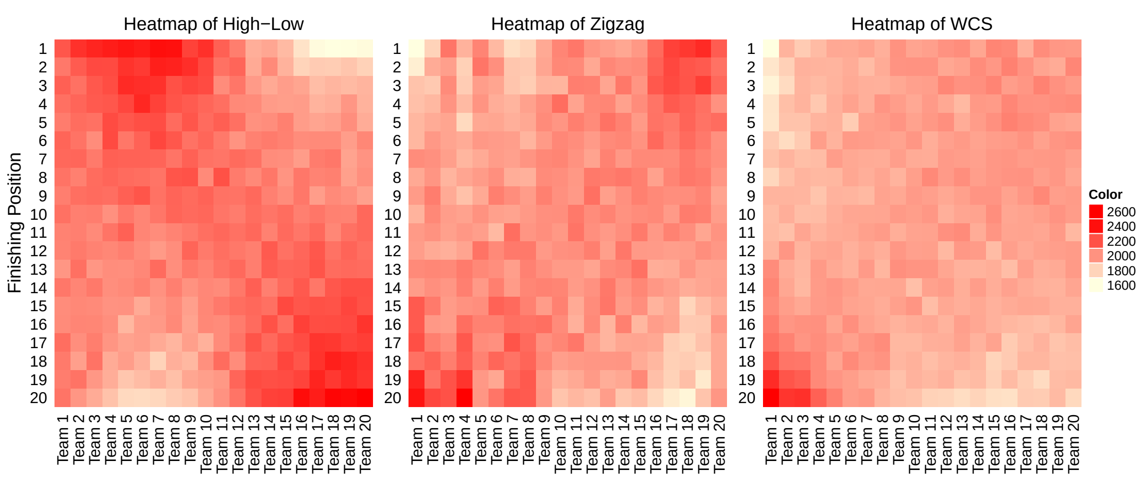

Figure 4.2: Heatmaps of the frequency tables produced by High-Low, Zigzag and WCS.

Chapter 5

Discussion

In most sports, it is difficult for players of varying skills to compete fairly against each other. For example, it is difficult to imagine an average person matched up against Usain Bolt in a 100m race.

However, in golf, a comprehensive handicap system has been devised to allow players of vastly different abilities to compete fairly against one another. Unfortunately, the handicap system can be far from fair in particular competitions, and there are many types of competitions in golf. For example, golf can be played according to match play or stroke play, golf can be played in two versus two settings or in tournament settings, and golf can be played in various formats such as best-ball, foursomes, aggregate, scrambles, etc.

In this project, we introduce a new method WCS to improve fairness where teams of sizes two and four are formed in net best-ball competitions. The method is supported by statistical theory, and simulation studies also suggest that WCS outperforms other standard team formation methods.

The key component of our proposal is the recognition of variability in golf performance. Our proposal is simple and straightforward to implement. Our method WCS is based on the heuristic of pairing the most variable golfers with the least variable golfers.Perhaps the variability aspect can be introduced to improve the fairness in other types of golf competitions.

Bibliography

Benincasa, G, Pavlikov, K., Hearn, Algorithms and Software for the Golf Director Problem. (2017). *Optimization Online*. Accessed online June 20, 2018 at

http://www.optimization-online.org/DB_FILE/2017/10/6299.pdf

Bingham, D.R. and Swartz, T.B. (2000). Equitable handicapping in golf. *The American Statistician*, 54(3), 170-177.

Golf Canada (2016). Golf Canada Handicap Manual. Accessed online August 21, 2017 at

http://golfcanada.ca/app/uploads/2016/02/2016-Handicap-Manual-6x9-ENG-FA-web.pdf

Grasman, S.E. and Thomas, B.W. (2013). Scrambled experts: team handicaps and win probabilities for golf scrambles. *Journal of Quantitative Analysis in Sports*, 9, 217-227.

Hurley, W.J. and Sauerbrei, T. (2015). Handicapping net best-ball team matches in golf. *Chance*, 28, 26-30.

Kupper, L.L., Hearne L.B., Martin, S.L. and Griffin, J.M. (2001). Is the USGA golf handicap system equitable? *Chance*, 14(1), 30-35.

Madras, D. (2017). Improving fairness in match play golf through enhanced handicap allocation. *Journal of Sports Analytics*, To appear.

Nadarajah, S. and Kotz, S. (2008). Exact distribution of the max/min of two gaussian random variables. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 16(2), 210-212.

Pavlikov, K., Hearn, D. and Uryasev, S. (2014). The golf director problem: forming teams for club golf competitions. In *Social Networks and the Economics of Sports*, P.M. Pardalos, and V. Zamaraev, editors, Springer International Publishing: Switzerland, 157-170.

Pollock, S.M. (1974). A model for the evaluation of golf handicapping. *Operations Research*, 22(5), 1040-1050.

Scheid, F. (1977). An evaluation of the handicap system of the United States Golf Association. In *Optimal Strategies in Sports*, S.P. Ladany and R.E. Machol, editors, North Holland: Amsterdam, 151-155.

Scheid, F. (1990). On the normality and independence of golf scores, with applications. In *Science and Golf: Proceedings of the First World Scientific Congress of Golf,* A.J. Cochran, editor, London:E. and F.N. Spon, 147-152.

Siegbahn, P. and Hearn, D. (2010). A study of fairness in fourball golf competition. In *Optimal Strategies in Sports Economics and Management*, S. Butenko, J. Gil-Lafuente and P.M. Pardalos, editors, Springer-Verlag: Heidelberg, 143-170.

Swartz, T.B. (2009). A new handicapping system for golf. *Journal of Quantitative Analysis in Sports*, 5(2), Article 9.

USGA (2016). USGA Handicap System Manual. Accessed online August 21, 2017 at http://www.usga.org/Handicapping/handicap-manual.html#!rule-14367

Wu, L., Chow-White, P. and Swartz, T.B. (2018). Net best-ball team composition in golf. *Submitted.*

Yun, H. (2011a). History of handicapping, part I: roots of the system. Accessed online August 21, 2017 at http://www.usga.org/articles/2011/10/history-of-handicapping-part-i–roots-of-the-system-21474843620.html

Yun, H. (2011b). History of handicapping, part II: increasing demand. Accessed online August 21, 2017 at http://www.usga.org/content/usga/home-page/articles/2011/10/history-of-handicapping-part-ii-increasing-demand-21474843658.html

Yun, H. (2011c). History of handicapping, part III: USGA leads the way. Accessed online August 21, 2017 at http://www.usga.org/content/usga/home-page/articles/2011/10/history-of-handicapping-part-iii-usga-leads-the-way-21474843686.html

Yun, H. (2011d). History of handicapping, part IV: the rise of the slope system. Accessed online August 21, 2017 at http://www.usga.org/content/usga/home-page/articles/2011/10/history-of-handicapping-part-iv-the-rise-of-the-slope-system-21474843751.html

Appendix A

Code

Below is the R code for the simulation study in Section 4.1. The code for the simulation study in Section 4.2 is more complex and is available from the author upon request.

```
##################################################################
#### Handicapping Best-Ball Events in Golf
##################################################################

############################
# Load libraries and read in data
############################
library(dplyr)
library(combinat)
library(stringr)
golf <- read.csv("../data/golf.csv", header = T)

############################
# Parameters definition
############################
tournament = 1000
no.of.team = 4
no.of.holes = 18
no.of.players = 8
seeds = 33

# select 8 players out of 16 in the original golf.csv
# pick 2nd, 4, 6, 8, 10, 12, 14, 16th
even_indexes <- seq(2, no.of.players*2 ,2)
golf <- golf[even_indexes,]
golf$group <- append(seq(1, no.of.players/2, 1), rev(seq(1, no.of.players/2, 1)))
golf$var <- (golf$sigma) ^ 2

# the dataframe golf is the same as demonstrated in Table 4.1

# calculate the variance of a two-man team
golf$team <- golf$group
pair.by.var <- golf %>% dplyr::group_by(team) %>%
  dplyr::summarize(total = sum(var)) %>%
  dplyr::ungroup() %>%
  dplyr::arrange(-total)
pair.by.var


##############################
####### Functions
##############################

```

```
DIM <- function(mat) {
  # function to calculate the no of rows in both single vector and matrix
  # since perform nrow() on a single vector will give NULL
  if (is.null(nrow(mat)))
  {
    size = 1
    return (size)
  }
  else
  {
    size = nrow(mat)
    return (size)
  }
}

sim_func <- function( Nsim, cutoff1, cutoff2, cutoff3, cutoff4)
{
  # function to generate net score probability distribution according to Table 4.1
  sim.results <- runif(n = Nsim, min = 0, max = 1)
  sim.hole_vec <- ifelse( sim.results < cutoff1, yes = -1,
                          no = ifelse(sim.results > cutoff1 & sim.results <= cutoff2,
                               yes = 0,
                                      no = ifelse(sim.results > cutoff2 & sim.results
                                           <= cutoff3, yes = 1,
                                                  no = 2
                                      )
                          )
  )
  return (sim.hole_vec)
}

# testing function
# table(sim_func( 100, cutoff1, cutoff2, cutoff3, cutoff4))


####################################
######### Generate All Possible Team Combinations
####################################

# calculate the total number of possible team formations (105 in total)
choose(n = 8, k = 2) * choose(n = 6, k = 2) * choose( n = 4, k = 2) / factorial(4)

labels <- c(1,2,3,4,5,6,7,8)
labels.comb <- do.call(cbind, combinat::permn(unique(labels)))
all_comb <- t(apply(labels.comb, 2, function(x) x[labels]))
kk <- all_comb
new <- matrix(NA,nrow=nrow(all_comb),ncol=8)

# swap the number/golfer index from low to high within team
new[,1] <- ifelse(all_comb[,1] < all_comb[,2], all_comb[,1], all_comb[,2] )
new[,2] <- ifelse(all_comb[,1] < all_comb[,2], all_comb[,2], all_comb[,1] )

new[,3] <- ifelse(all_comb[,3] < all_comb[,4], all_comb[,3], all_comb[,4] )
new[,4] <- ifelse(all_comb[,3] < all_comb[,4], all_comb[,4], all_comb[,3] )

new[,5] <- ifelse(all_comb[,5] < all_comb[,6], all_comb[,5], all_comb[,6] )
new[,6] <- ifelse(all_comb[,5] < all_comb[,6], all_comb[,6], all_comb[,5] )

new[,7] <- ifelse(all_comb[,7] < all_comb[,8], all_comb[,7], all_comb[,8] )
new[,8] <- ifelse(all_comb[,7] < all_comb[,8], all_comb[,8], all_comb[,7] )
#dim(new)

# remove duplicates
rm.dup <- unique(new) %>% as.data.frame(.)

rm.dup$x1 <- apply( rm.dup[ ,c(1,2) ] , 1 , paste , collapse = "" )
rm.dup$x2 <- apply( rm.dup[ ,c(3,4) ] , 1 , paste , collapse = "" )
```

```
rm.dup$x3 <- apply( rm.dup[ ,c(5,6) ] , 1 , paste , collapse = "" )
rm.dup$x4 <- apply( rm.dup[ ,c(7,8) ] , 1 , paste , collapse = "" )
rm.dup <- rm.dup %>% dplyr::select(x1:x4)

# step 1: sort the data row-wise in ascending order
# step 2: convert the matrix to dataframe but its type is characters
# step 3: convert the type from character to numeric
# step 4: find unique rows
# step 5: check the rows = 105
AllComb.df <- t(apply(rm.dup, 1,FUN=function(x) sort(x, decreasing=F)))  %>%
  as.data.frame( ., stringsAsFactors = F) %>%
  dplyr::mutate_if(is.character,as.numeric) %>% unique()

dim(AllComb.df) # this should be equal to 105 different team formations



####################################
######### simulation
####################################
ptm <- proc.time()
set.seed(seeds)
cum.score <- matrix(data = NA, nrow = tournament, ncol = no.of.team*105)

for (kk in 1:tournament)
{
  sim.hole <- matrix(data = NA, nrow = no.of.players, ncol = no.of.holes)

  for (gg in 1:nrow(golf))
  {
    # determine the cutoffs for the input of sim.hole uniform distribution
    cutoff1 <- cumsum(as.numeric(golf[gg ,c("minus1", "zero", "plus1", "plus2")]))[1]
    cutoff2 <- cumsum(as.numeric(golf[gg ,c("minus1", "zero", "plus1", "plus2")]))[2]
    cutoff3 <- cumsum(as.numeric(golf[gg ,c("minus1", "zero", "plus1", "plus2")]))[3]
    cutoff4 <- cumsum(as.numeric(golf[gg ,c("minus1", "zero", "plus1", "plus2")]))[4]

    # generate 18 holes for one player
    sim.hole[gg,] <- sim_func( Nsim = no.of.holes, cutoff1, cutoff2, cutoff3, cutoff4
        )
  }


  for ( tt in 1:nrow(AllComb.df))
  {
    # determine team assigned for group_by team function later
    # team 1
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 1], start = 1, end = 1))] <- 1
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 1], start = 2, end = 2))] <- 1

    # team 2
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 2], start = 1, end = 1))] <- 2
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 2], start = 2, end = 2))] <- 2

    # team 3
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 3], start = 1, end = 1))] <- 3
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 3], start = 2, end = 2))] <- 3

    # team 4
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 4], start = 1, end = 1))] <- 4
    golf$team[ as.numeric( str_sub( AllComb.df[ tt, 4], start = 2, end = 2))] <- 4

    sim.df <- data.frame(golf, sim.hole)

    # find the minimum score for all 18 holes within the same team
    best.ball <- sim.df %>%
      dplyr::group_by(team) %>%
      dplyr::summarise_at(.cols = vars(starts_with("X")),
```

```
                    .funs = c(minimum = "min")) %>%
      as.data.frame() %>%
      dplyr::select(X1_minimum:X18_minimum) %>% t()

    #########################################################
    # calculate the sum of scores over all 18 holes
    # 4 * 105 different team assignment = 620 columns in the matrix
    cum.score[ kk, (4*(tt - 1 ) + 1): (4*tt) ] <- apply(best.ball, MARGIN = 2, sum)
  }

}

# save(cum.score, file = "cum.score_2man.Rdata")
load("cum.score_2man.Rdata")
temp2 <- matrix(data = 0, nrow = tournament, ncol = 4*105)
for (aa in 1:tournament)
{
  for (bb in 1:105)
  {
    # for stroke play, cum.scores are the total number of scores
    temp2[aa,(4*(bb - 1 ) + 1): (4*bb)] <- rank(cum.score[aa,(4*(bb - 1 ) + 1): (4*bb
        ) ], ties.method = c("random"))
  }
}
(time <- proc.time() - ptm)
time
# user  system elapsed
# 609.244   8.127 623.583



####################################
######### Produce the table for finishing positions (Table 4.2)
####################################

RankTable <- apply(temp2, 2, function(x) table(x) )

# Define a function to determine the finishing positions
get_position <- function (x,nth_place){
  if (sum(c(1:4) %in% as.numeric(names(RankTable[[x]])))==4)
  {
    res <- as.numeric(RankTable[[x]][nth_place])
  } else {
    if (nth_place==1)
    {
      res <- 0
    }
    else if ((nth_place==2)|(nth_place==3)|(nth_place==4) ){
      res <-  as.numeric(RankTable[[x]][(which(names(RankTable[[x]])==nth_place))])
    }
  }
  return(res)
}


first_pos <- matrix(unlist(lapply( c(1:420), get_position,nth_place=1)), nrow = 105,
    ncol = 4, byrow=T)
second_pos <- matrix(unlist(lapply( c(1:420), get_position,nth_place=2)), nrow = 105,
     ncol = 4, byrow=T)
third_pos <- matrix(unlist(lapply( c(1:420), get_position,nth_place=3)), nrow = 105,
    ncol = 4, byrow=T)
fourth_pos <- matrix(unlist(lapply( c(1:420), get_position,nth_place=4)), nrow = 105,
     ncol = 4, byrow=T)

all.place.mat <- cbind(first_pos,second_pos,third_pos,fourth_pos)
# colnames(first_pos) <- c("WinT1", "WinT2", "WinT3", "WinT4")
chi.sq <- apply(((all.place.mat - 250)^2)/250,1,sum)
```

```
All.dist.final <- cbind(AllComb.df,chi.sq,all.place.mat) %>% dplyr::arrange(chi.sq)

# matrix(All.dist.final[1,c(6:21)],nrow=4,byrow = T)

dist.showing <- All.dist.final %>% dplyr::mutate (
  team1 = paste( str_sub( All.dist.final[ , 1], start = 1, end = 1),
                 str_sub( All.dist.final[ , 1], start = 2, end = 2), sep = " & "),
  team2 = paste( str_sub( All.dist.final[ , 2], start = 1, end = 1),
                 str_sub( All.dist.final[ , 2], start = 2, end = 2), sep = " & "),
  team3 = paste( str_sub( All.dist.final[ , 3], start = 1, end = 1),
                 str_sub( All.dist.final[ , 3], start = 2, end = 2), sep = " & "),
  team4 = paste( str_sub( All.dist.final[ , 4], start = 1, end = 1),
                 str_sub( All.dist.final[ , 4], start = 2, end = 2), sep = " & ")
) %>% dplyr::select(team1, team2, team3, team4, chi.sq)

# write.csv(dist.showing,file="2man_final_table.csv",row.names = F)
```

2ManTeam.R