No-Ad Game scoring within a Best of 5 Sets Tennis Matches

By
Tristan Barnett, Graham Pollard, Alan Brown, Geoff Pollard and Vladimir Ejov

Abstract
Tennis matches that take much longer than expected are a problem in several ways. They can delay the starting time of the following match, cause issues for broadcasters, lead to an increased number of injuries, and decrease the winner's chance of winning in the next round. In this paper several alternative game structures for possible use in reducing the length of best-of-5 set matches are studied. Also, criteria for comparing two or more tennis match scoring systems are outlined.

Keywords: alternative game scoring systems for tennis, long matches in tennis, mean duration, variability of duration of tennis matches, efficiency of tennis scoring, changes to tennis scoring, parameter values in tennis.

Introduction

The uncertain and highly variable length of games, sets and matches in tennis has been a concern for players, television, spectators, as well as tournament directors. It remains a concern.

Some matches have been observed to last more than 5 hours even though the 5^{th} set was not a 'long' advantage set. For example, in the 2012 Australian Open Final, Djokovic beat Nadal, 5-7, 6-4, 6-2, 6-7, 7-5 in 5 hours and 53 minutes. Whilst long matches can be exciting as a stand-alone match (as in a final), it can be seen as unfair in the tournament setting as the winner can be too exhausted to do justice to his performance in the next round. This typically can occur in men's grand slam singles matches as they play best-of-5 set matches. Thus, it would be useful to have a scoring system that reduces the likelihood of such long matches, whilst keeping other match characteristics (such as the probability of the stronger player winning) much the same as they are at present.

Over the last several decades a considerable amount of research has been carried out on the match characteristics of various tennis scoring systems. When considering alternative scoring systems, it is not sufficient to consider just the mean and variance of the duration of a match. Issues related to the skewness of this distribution of duration are an important consideration. The probability that the better player wins also needs to be considered. Such measures or characteristics are available and are important in deciding whether a scoring system is acceptable or not. It is noted that there is typically a need for compromise when considering two or more scoring systems, as it is unlikely that one system is best on all such measures.

The major purpose of this research is to achieve a greater understanding of the characteristics of several match scoring systems using variations of the No-Ad game concept, which was described and studied by Pollard and Noble (2004) and is currently used in doubles (excluding grand slams where a standard deuce game is used). In the No-Ad game a player needs to win 4 points in order to win the game (and if the score line reaches 3 points-all in that game, the player who wins the next point wins that game). At most 7 points are played in the No-Ad game and this characteristic helps to reduce the skewness of the distribution of duration of a set and match of tennis. Some other game scoring systems are also considered in this paper.

The idea of "deuce" was introduced (at least as far back as 1490) for a simple reason … to ensure that the game could not be won by a one-point difference in the players' scores. Deuce was derived from the French "a deux du jeu"… two points away from game. It is reasonable to believe that there was no mathematics carried out on this deuce game back in 1490 concerning how it would affect the game of tennis into the future (Barnett, 2012).

Pollard and Noble (ibid) studied the efficiency and some duration characteristics of a best of 3 sets match using the No-Ad game. They also set up the 50-40 game in which server needs to win 4 points in a game whilst the receiver needs to win only 3 points to win the game. The concept behind such a game structure was that the server has the advantage of serving within the game but has the disadvantage of needing to win one more point than the receiver in order to win. They noted, for example, the increased efficiency of best of 3 set tennis scoring when used for 'strong' servers as in men's doubles but did not study the best of 5 set matches, which are considered in this paper.

Recently, Pollard and Barnett (2018) reported on the 50-40 game and a few variations of it *within a single set of tennis*. One variation they studied was the 50-40, 40-0, 40-15 game which is a 50-40 game modified so that the server wins the game if the score reaches 40-0 or 40-15. The logic behind a scoring system such as this is that there is little point (in terms of efficiency and duration) in playing points that are relatively unimportant and unexciting. Further, in their discussion section they suggested a couple of further modifications of the 50-40 game that could be usefully studied. In their work they considered just a single set of tennis and whilst these single set results give some useful insights into the likely characteristics of a best of 3 or best of 5 sets match when using such games, the study of a complete match gives greater clarity regarding any preferred system. This is done in this paper.

Alternative game structures to address the problem of 'long' best of 5 set matches

In this study we consider the **best of 5 tiebreak sets** using the two official game scoring systems of the Rules of tennis (advantage/deuce games and no ad games) as well as three possible alternatives.

(1) Advantage/Deuce games – a player needs to win 4 points but if the score line reaches 3 points-all, then a player must be 2 points ahead to win the game.

(2) No-Ad – a player needs to win 4 points in order to win the game. If the score line reaches 3 points-all, the player who wins the next point wins the game. At most 7 points are played in the No-Ad game.

(3) No-Ad* - a player needs to win 4 points but if the score line reaches deuce, then a player must win 2 more points to win the game. At most 9 points are played in the No-Ad* game.

(4) 50-40 (as defined by Pollard and Noble (ibid)) – the server needs to win 4 points whilst the receiver needs to win just 3 points in order to win the game. At most 6 points are played in this type of game.

(5) 50-40* - server needs to win 4 points and receiver needs to win just 3 points but if the score line reaches 3-2 (40-30) then the player who wins two more points wins the game. At most 8 points are played in this game.

Criteria for comparing tennis scoring systems

In this study of the best of 5 tiebreak sets matches for men, where every game is one of the above-mentioned game scoring systems, the match characteristics of interest are

(1) Probability that the stronger player wins, P
(2) Expected value of number of points played (duration) in the match, E(D)
(3) Standard deviation of the number of points played in the match, SD(D)

(4) Efficiency of the scoring system
(5) Coefficient of skewness of the number of points played in the match
(6) The 95%, 99% and 99.5% points in the cumulative distribution of duration, denoted by CD95, CD99 and CD99.5.

Note that the efficiency of the tennis scoring system was devised in a very elegant paper by Roger Miles (Miles, 1984). The efficiency of a tennis scoring system with key characteristics P, the probability that the better player wins, and m, the mean duration (mean number of points played in the match), is equal to:

$$(2(P-Q)\ \ln(P/Q))\ /\ (m\ (p_A-p_B)\ \ln(p_A q_B/p_B q_A))$$

where $Q = 1-P$, p_A is the probability player A wins a point on service, p_B is the probability player B wins a point on service, $q_A = 1-p_A$ and $q_B = 1-p_B$.

Given two scoring systems with the same mean duration, the one in which the better player A has a higher probability of winning has the greater efficiency. Correspondingly, given two scoring systems with the same likelihood of the better player winning, the one that has the smaller mean duration has the greater efficiency. Note that the efficiencies of tennis scoring systems are typically a lot less than 1 mainly because of the nested nature (points, games, sets) of tennis scoring using 'best of' structures. Very efficient scoring systems do not have 'best of' structures. They also have very large variances of duration, and this makes them quite inappropriate for scoring in tennis.

Note that the characteristics CD95, CD99, and CD99.5 should be sufficient for comparing the upper tails of the duration distributions, as (only) 127 five-set matches are played each year in each Grand Slam Men's Singles event.

Parameter values

The key input parameters in modelling a men's singles tennis match between player A and player B are p_A = probability that player A wins a point on his serve, and p_B = probability that player B wins a point on his serve.

Cross and Pollard (2011) noted that for men singles at the four Grand Slam events in 2008, the proportion of points won on service averaged 0.631, 0.621, 0.667 and 0.643 at Australian Open, French Open, Wimbledon and US Open respectively. They reported that these values 'had not changed much over the years [1999 to 2009]', except for the French Open ('associated with a considerable increase in first service speed' (Cross and Pollard (2009))). As these have an average value of 0.64, this value is used as the most appropriate average value for this study.

In the Cross and Pollard study the proportion of points won on service by the winner minus the proportion won on service by the loser was 0.11. This figure is biased in favour of the winner. For example, using simulation methods in a study of bias in sporting statistics, Pollard et al (2010) noted…"As the winner must have won the last point, last game and last set, the winner's service statistics can have an upwards bias, and the loser's service statistics a downwards bias.", and…"In the best of three tiebreak sets match between two *equal* players (with $p_A = p_B = 0.65$), the proportion of points won on service by the eventual winner is shown to be about 0.065 on average greater than the proportion of points won on service by the loser. For a best of five tiebreak sets match between these two equal players, this difference is shown to average about 0.049."

It is important that such biases are considered when working with reported statistics. We have done this in deciding the parameters to use in our study.

Taking 0.04 as a reasonable difference between the serving p-values in a 'moderately close' match, appropriate values for the parameters in a typical or average men's singles match are $p_A = 0.66$, $p_B = 0.62$.

It is noted here that, in an article using data from the 2016 Rio Olympics, Carl Bialik (2016) concluded that the service success rates for men's singles was 63%. It is noted that this percentage is quite similar to the figure above, and to the assumptions we make in our modelling.

It was anticipated that whilst no scoring system would be 'best' with respect to all characteristics, one scoring system might in some sense be best "overall'.

Method of Analysis

Most of these results are numerically exact and were developed using recursive formulas in an Excel spreadsheet (Barnett, 2016). The theory behind the recursive formulas is now outlined.

Our use of the moment generating functions to add up the statistics on the number of points is based on the following theorem.

Theorem 1. Let $Z = X + Y$, where X and Y are independent random variables, then $M_Z(t) = M_X(t)M_Y(t)$.

Proof. $M_Z(t) = E(e^{Zt}) = E(e^{(X+Y)t}) = E(e^{Xt})E(e^{Yt}) = M_X(t)M_Y(t)$.

The algorithm for extracting the moments of the sum of independent random variables is the following:

$m_{1Z} = m_{1Y} + m_{1X}$

$m_{2Z} = m_{2Y} + 2m_{1X}m_{1Y} + m_{2X}$

$m_{3Z} = m_{3Y} + 3m_{1X}m_{2Y} + 3m_{2X}m_{1Y} + m_{3X}$

$m_{4Z} = m_{4Y} + 4m_{1X}m_{2Y} + 6m_{2X}m_{2Y} + 4m_{3X}m_{1Y} + m_{4X}$

This is a simple corollary of Theorem 1, which can be obtained by successive differentiation with respect to t, and putting t to 0.

The score for a match in progress will be denoted by (a, b : c, d : e, f), where (a, b) is the score in points, (c, d) is the score in games, and (e, f) is the score in sets, for player A and player B respectively. We will use a truncated form of this notation whenever it is convenient so to do.

A tennis match consists of four levels - (points, games, sets, match). It becomes necessary to represent:
points in a point as pp,
points in a game as pg,
points in a tiebreak game as pg_T,
points in a tiebreak set as ps_T
points in a best of 5 all tiebreak set match as pm_{5T}.

Let s_A, s_B represent the condition that player A and player B, respectively served first at the beginning of a set. Let c_A, c_B represent the condition that player A and player B, respectively are currently serving in the set at the score (a, b : c, d). If (a, b) is not a boundary score for the current game then
$s_A = c_A$ and $s_B = c_B$, if $(c + d) \bmod 2=0$

$s_A = c_B$ and $s_B = c_A$, if (c + d) mod 2=1
except in the case of the tiebreak game of the tiebreak set, with c = 6, d = 6, when
$s_A = c_A$ and $s_B = c_B$, if (a + b) mod 4=0 or 3
$s_A = c_A$ and $s_B = c_B$, if (a + b) mod 4=1 or 2

Let $P^{psT}(a, b : c, d|s_A)$ represent the probability of player A winning a tiebreak set at this score, and player A serving first in the current set. Let $Y^{psT}(a, b : c, d|s_A)$ be the number of points remaining in the set at this score with player A serving first in the current set. This number is a random variable. Let $M_{YpsT(a,b:c,d|sA)}(t)$ be its moment generating function.

Similarly let $P^{psT}(a, b : c, d|w_A, s_A)$ represent the probability of player A winning a tiebreak set at this score, and player A serving first in the current set. Let $Y^{psT}(a, b : c, d|w_A, s_A)$ be a random variable of the number of points remaining in the set at this score conditional on player A both winning the set, and serving first in the current set. Let $M_{YpsT(a,b:c,d|wA,sA)}(t)$ be its moment generating function conditional on player A both winning the set, and serving first in the current set.

Many variants of this notation will be used. The representation of the score will be restricted whenever it is not essential to display the full score. Other symbols include B for player B, l for the condition of losing, and n for the condition of serving next.

The next step is to introduce weighted moment generating functions. Let X be a conditional random variable. Let C be the condition that X occurs with probability p_X.

Then
$W_{X|C}(t) = p_X M_X(t)$

This product of a probability and its associated moment generating function is defined as a weighted moment generating function. The weight is the probability measure such that the conditions applied to the random variable are true.

Denote by w_{nX} the weighted n^{th} moment of the random variable X. Then
$w_{nX} = p_X m_{nX}$ for n = 1, 2, 3, 4, ...

The more important situation for us arises when the score does change. Let X and Y be independent random variable with conditional probabilities p_X and p_Y, respectively, of occurring. Let Z denote the random variable for their sum, Z = X + Y when both X and Y occur. Then $p_Z = p_X p_Y$. It follows from Theorem 1 that the weighted moment generating functions satisfy
$W_{Z|C1,C2}(t) = W_{X|C1}(t) W_{Y|C2}(t)$.

The algorithm for extracting the weighted moments is the following:
$p_Z = p_X p_Y$
$w_{1Z} = p_X w_{1Y} + w_{1X} p_Y$
$w_{2Z} = p_X w_{2Y} + 2w_{1X} w_{1Y} + w_{2X} p_Y$
$w_{3Z} = p_X w_{3Y} + 3w_{1X} w_{2Y} + 3w_{2X} w_{1Y} + w_{3X} p_Y$
$w_{4Z} = p_X w_{4Y} + 4w_{1X} w_{2Y} + 6w_{2X} w_{2Y} + 4w_{3X} w_{1Y} + w_{4X} p_Y$

We now develop the algebra for weighted moment generating functions. We are able to add together two weighted moment generating functions whenever we encounter two mutually exclusive cases. Two simple examples where the score does not change are:
(a) Condition on initial server
$W_{YpsT(a,b:c,d)}(t) = W_{YpsT(a,b:c,d|sA)}(t) + W_{YpsT(a,b:c,d|sB)}(t)$

(b) Condition on winning or losing
$W_{YpsT(a,b:c,d|sA)}(t) = WY_{psT(a,b:c,d|wA,sA)}(t) + W_{YpsT(a,b:c,d|lA,sA)}(t)$

We now apply these ideas to the playing of a single point. In this case some of the notation appears to degenerate, so we must be careful. However, this analysis will be used whenever the score changes as a point is played in a game, a set, or a match.

Each point played is a single point, irrespective of the score. For player A serving, the probability of winning the point is denoted by p_A irrespective of the score and $q_A = 1-p_A$. Let $P^{pp}(()|c_A, w_A)$ and $P^{pp}(()|c_A, l_A)$ represent the probabilities of player A winning and losing a point on serve respectively from score line () within the point. It follows that:
$P^{pp}(()|c_A, w_A) = p_A$
$P^{pp}(()|c_A, l_A) = q_A$

Let $Y^{pp}(()|c_A)$ represent the number of points remaining in the point from score line () with player A serving. Each point played is a single point, so $Y^{pp}(()|c_A) = 1$. Let $Y^{pp}(()|c_A,w_A)$ and $Y^{pp}(()|c_A,l_A)$ represent the number of points remaining in the point from score line () given player A won and lost the point respectively with player A serving.

Therefore:
$M_{Ypp(()|cA)}(t) = E(e^{Ypp(()|cA)t}) = E(e^t) = e^t$
$W_{Y\,pp(()|cA,wA)}(t) = P^{pp}(()|c_A,w_A)M_{Y\,pp(()|cA)}(t) = p_A e^t$

This is a fundamental brick in the model.

It is easy to check that
$w_n(Y^{pp}(()|c_A, w_A)) = p_A$ for n = 0, 1, 2, 3, 4, ...

Likewise $W_{Ypp(()|cA,lA)}(t) = P^{pp}(()|c_A, l_A)M_{Ypp(()|cA)}(t) = q_A E(e^t)$

and
$w_n(Y^{pp}(()|c_A, l_A)) = q_A$ for n = 0, 1, 2, 3, 4, ...

a) Number of points in a game

Let $W_{Ypg(a,b|cA,wA)}(t)$ and $W_{Ypg(a,b|cA,lA)}(t)$ represent the weighted moment generating functions of the number of points remaining in the game from score line (a, b) given player A is serving and player A won and lost the game respectively.

Theorem 2. $W_{Ypg(a,b|cA,wA)}(t)$
$= W_{Ypp(()|cA,wA)}(t)W_{Ypg(a+1,b|cA,wA)}(t) + W_{Ypp(()|cA,lA)}(t)W_{Ypg(a,b+1|cA,wA)}(t)$

Proof. $M_{Ypg(a,b|cA,wA)}(t) = E(e^{tY\,pg(a,b|cA,wA)})$ is an expectation that is calculated before the point at score (a, b) has been played. The point played is won and lost with probability p_A and q_A respectively, where $p_A+q_A=1$ since there are only two possible outcomes. When we try to recalculate the original expectation after the point has been played, we obtain the weighted sum of two expressions

$M_{Ypg(a,b|cA,wA)}(t)$
$= p_A E(e^{t(1+Y\,pg(a+1,b|cA,wA))})P^{pg}(a + 1, b|c_A, w_A)/P^{pg}(a, b|c_A, w_A)$

$+q_A E(e^{t(1+Y\ pg(a,b+1|cA,wA))}) P^{pg}(a, b + 1|c_A, w_A)/P^{pg}(a, b|c_A, w_A)$

where the odds ratios
$P^{pg}(a + 1, b|c_A, w_A)/P^{pg}(a, b|c_A, w_A)$ and $P^{pg}(a, b + 1|c_A, w_A)/P^{pg}(a, b|c_A, w_A)$ reflect the changes in the chances of player A winning when the score is updated after winning or losing the point, respectively. The count of 1 for the point played is independent of the distribution of the remaining points after the point has been played, so, as in
Theorem 1, we can factorize the expectations to obtain
$E(e^{t(1+Y\ pg(a+1,b|cA,wA))}) = E(e^t)E(e^{t(Y\ pg(a+1,b|cA,wA))})$ and
$E(e^{\ t(1+Y\ pg(a,b+1|cA,wA))}) = E(e^t)E(e^{t(Y\ pg(a,b+1|cA,wA))})$.

After some rearrangement we find that
$P^{pg}(a, b|c_A, w_A)M_{Ypg(a,b|cA,wA)}(t)$
$= p_A E(e^t)P^{pg}(a + 1, b|c_A, w_A)E(e^{t(Y\ pg(a+1,b|cA,wA))})$
$+q_A E(e^t)P^{pg}(a, b + 1|c_A, w_A)E(e^{t(Y\ pg(a,b+1|cA,wA))})$
The only step that is left is the identification of the various terms in this expression as weighted moment generating functions, to obtain
$W_{Ypg(a,b|cA,wA)}(t)=W_{Ypp(()|cA,wA)}(t)W_{Ypg(a+1,b|cA,wA)}(t)+W_{Ypp(()|cA,lA)}(t)W_{Ypg(a,b+1|cA,wA)}(t)$

Note carefully in Theorem 2 how first we are able to multiply the weighted moment generating functions on each path of this branching process which arises when scoring, because the steps on each branch are independent; and then add the results of this multiplication, because the paths are mutually exclusive.

It follows that $W_{Ypg(a,b|cA,wA)}(t) = P^{pg}(a, b|c_A, w_A)M_{Y\ pg(a,b|cA,wA)}(t)$
where $M_{Ypg(a,b|cA,wA)}(t)$ is the moment generating function of the random variable $Y^{pg}(a, b|c_A, w_A)$.

By successive differentiation with respect to t from Theorem 2, and setting t = 0 we obtain the following recurrence formulas.

$w_1(Y^{pg}(a, b|c_A, w_A)) = p_A w_1(Y^{pg}(a+ 1, b|c_A, w_A)) +q_A w_1(Y^{pg}(a, b+ 1|c_A, w_A)) +p_A P^{pg}(a+ 1, b|c_A, w_A) + q_A P^{\ pg}(a, b + 1|c_A, w_A)$

$w_2(Y^{pg}(a, b|c_A, w_A)) = p_A w_2(Y^{pg}(a+1, b|c_A, w_A))+q_A w_2(Y^{pg}(a, b+1|c_A, w_A))+2p_A w_1(Y^{pg}(a+ 1, b|c_A, w_A)) + 2q_A w_1(Y^{pg}(a, b + 1|c_A, w_A)) + p_A P^{pg}(a + 1, b|c_A, w_A) + q_A P^{pg}(a, b + 1|c_A, w_A)$

$w_3(Y^{pg}(a, b|c_A, w_A)) = p_A w_3(Y^{pg}(a+1, b|c_A, w_A))+q_A w_3(Y^{pg}(a, b+1|c_A, w_A))+3p_A w_2(Y^{pg}(a+ 1, b|c_A, w_A)) + 3q_A w_2(Y^{pg}(a, b+ 1|c_A, w_A))+3p_A w_1(Y^{pg}(a+ 1, b|c_A, w_A))+3q_A w_1(Y^{pg}(a, b+ 1|c_A, w_A)) + p_A P^{pg}(a + 1, b|c_A, w_A) + q_A P^{pg}(a, b + 1|c_A, w_A)$

$w_4(Y^{pg}(a, b|c_A, w_A)) = p_A w_4(Y^{pg}(a+1, b|c_A, w_A))+q_A w_4(Y^{pg}(a, b+1|c_A, w_A))+4p_A w_3(Y^{pg}(a+ 1, b|c_A, w_A)) + 4q_A w_3(Y^{pg}(a, b+ 1|c_A, w_A)) + 6p_A w_2(Y^{pg}(a+ 1, b|c_A, w_A)) + 6q_A w_2(Y^{pg}(a, b+ 1|c_A, w_A))+4p_A w_1(Y^{pg}(a+1, b|c_A, w_A))+4q_A w_1(Y^{pg}(a, b+1|c_A, w_A))+p_A P^{pg}(a+1, b|c_A, w_A)+ q_A P^{pg}(a, b + 1|c_A, w_A)$

Boundary Values:
$w_n(Y^{pg}(a, b|c_A, w_A)) = 0$, if a = 4 and $0 \le b \le 2$; b = 4 and $0 \le a \le 2$
$w_1(Y^{pg}(3, 3|c_A, w_A)) = 2p^2_A/(2p^2_A-2p_A+1)^2$
$w_2(Y^{pg}(3, 3|c_A, w_A)) = 4p^2_A(1-2p^2_A+2p_A) / (2p^2_A-2p_A+1)^3$
$w_3(Y^{pg}(3, 3|c_A, w_A)) = 8p^2_A(4p^4_A-8p^3_A-4p^2_A+8p_A+1)/(2p^2_A-2p_A+1)^4$
$w_4(Y^{pg}(3, 3|c_A, w_A)) = 16p^2_A(1-2p^2_A+2p_A)(4p^4_A-8p^3_A-16p^2_A+20p_A+1)/(2p^2_A-2p_A+1)^5$

Similar recursion formulas with boundary values can be obtained for $w_n(Y^{pg}(a, b|c_A, l_A))$.

Let $M_{Ypg(a,b|cA)}(t)$ represent the moment generating function of the number of points remaining in a game at point score (a, b) for player A serving.

Using the rule for combining weighted moment generating functions with mutually exclusive conditions we obtain $M_{Ypg(a,b|cA)}(t) = W_{Ypg(a,b|cA,wA)}(t) + W_{Ypg(a,b|cA,lA)}(t)$ since the probability that a game will eventually end is 1.

Converting moments to parameters of distribution (mean, variance, coefficients of skewness and excess kurtosis) can readily be obtained.

Similar formulas and parameters of distribution can be obtained for when player B is serving such that $W_{Ypg(a,b|cB,wB)}(t)$ and $W_{Ypg(a,b|cB,lB)}(t)$ represent the weighted moment generating functions of the number of points remaining in the game from score line (a, b) given player B is serving and player B wins and loses the game respectively.

b) Number of points in a tiebreak game

The analysis of a tiebreak game is similar to that of a standard game except that it is necessary to allow for the rotation of service before each odd point in the tiebreak game.

c) Number of points in a tiebreak set

We study here the model for a tiebreak set. To account for the rotation of service in this type of set it is necessary to allow for the rotation of server at the beginning of each game. Using this convention, whenever a tiebreak game is required to resolve the winner of the set, this tiebreak game is marked to the server of the first point of the game, and hence to the server of the first point of the set when it comes to determining the first server of the next set. This rule applies irrespective of the outcome of the tiebreak game.

For player A serving in the first game of the set there are four cases to be dealt with separately. Consider the case where player A not only serves in the first game of the set, but wins the set, and serves in the first game of the next set.

Let $W_{YpsT\,(0,0:c,d|sA,wA,nA)}(t)$ represent the weighted moment generating function of the number of points remaining in a tiebreak set at point and game score (0, 0 : c, d) given player A served first, wins the set and is serving first in the next set to be played. Then by considering a complete game being played at that score we obtain, for c + d < 12

$W_{YpsT(0,0:c,d|sA,wA,nA)}(t)=W_{Ypg(0,0|cA,wA)}(t)W_{YpsT(0,0:c+1,d|sA,wA,nA)}(t)+W_{Ypg(0,0|cA,lA)}(t)W_{YpsT(0,0:c,d+1|sA,wA,nA)}(t)$, for (c + d) mod 2=0

$W_{YpsT(0,0:c,d|sA,wA,nA)}(t)=W_{Ypg(0,0|cB,lB)}(t)W_{YpsT(0,0:c+1,d|sA,wA,nA)}(t)+W_{Ypg(0,0|cB,wB)}(t)W_{YpsT(0,0:c,d+1|sA,wA,nA)}(t)$, for (c + d) mod 2=1

There is a special case for the tiebreak game, with c = 6, d = 6, where due to the rotation of serve player A cannot serve first in the next set, so
$W_{YpsT\,(0,0:6,6|sA,wA,nA)}(t) = 0$, which simplifies to $W_{YpgT\,(0,0|cA,wA)}(t)=0$

d) Number of points in a best of 5 all tiebreak set match

Because we have to take into account both the winner of the current set and the server at the start of the next set, the recurrence formulas have to allow for four-way branching rather than the two-way branching that we have previously met.

For player A winning the match and currently serving,

$$W_{Ypm5T(0,0:0,0:e,f|cA,wA)}(t)=W_{YpsT(0,0:0,0|sA,wA,nA)}(t)W_{Ypm5T(0,0:0,0:e+1,f|cA,wA)}(t)+W_{YpsT(0,0:0,0|sA,lA,nA)}(t)W_{Ypm5T(0,0:0,0:e,f+1|cA,wA)}(t)+W_{YpsT(0,0:0,0|sA,wA,nB)}(t)W_{Ypm5T(0,0:0,0:e+1,f|cB,wA)}(t)+W_{YpsT(0,0:0,0|sA,lA,nB)}(t)W_{Ypm5T(0,0:0,0:e,f+1|cB,wA)}(t)$$

It has been demonstrated throughout this section that the total number of points played in a tennis match has a discrete distribution. The moments of this distribution can be calculated using a lattice model with the Markov property and a few other modest assumptions. The Normal distribution has been widely studied, and tables of the probabilities for this distribution are readily available. The basic idea of the Normal Power approximation is to use these tables to estimate the tail probabilities of other distributions. This method uses the first four moments and produces a continuous approximation to the cumulative distribution.

The Normal Power approximation has a weakness in that it can fail when fitting distributions that are multimodal. Therefore, special steps must be taken when estimating the distribution of points in a tennis match. The key observation is that the distribution of points in a tie-breaker set is unimodal. The Normal Power approximation can be safely used to estimate this distribution. The quality of this estimate can be checked using simulation. The distribution of points in 3 set endings, 4 set endings and 5 set endings are each unimodal in a best of 5 set all tie-breaker match, as they inherit the properties of a single tie-breaker set. Each of these distributions can be estimated, and the distribution of points for the complete match can be obtained by weighed addition, where the weights to be used are the probabilities of each type of ending.

Let X be a random variable with a cumulative distribution F(x), so that $P(X \geq x) = F(x)$

Let μ, σ, γ_1, γ_2 be the mean, standard deviation, skewness and excess kurtosis of X. Let Z be a standardized random variable with mean 0 and standard deviation 1, with

$P(Z \geq z) = P(X \geq x)$

Denote the cumulative Normal distribution by $\phi(.)$. Then the Normal Power approximation can be written as

$F(x) \approx \phi(y)$

with
$z = (x-\mu)/\sigma$

and
$y = z - 1/6\ \gamma_1(z^2 - 1) - 1/24\ \gamma_2(z^3 - 3z) + 1/36\ \gamma_1^2\ (4z^3 - 7z)$

Results

1. Firstly, we consider an 'average' or 'typical' men's singles match with p_A = 0.66 and p_B = 0.62. These (p_A, p_B) parameters represent an 'average' match in Grand Slam men's tennis and are particularly relevant for the US Open or Australian Open. The results for such a match are given in Table 1 for player A serving first in the match. Columns 2-6 in the tables are exact results (from the methodology). They were checked against the equivalent (exact) best of 3 tiebreak sets results in Pollard (1983).

B5 sets	P(A wins)	Mean	Efficiency	Stand Dev	Skew	CD95	CD99	CD99.5
Ad games	0.734	261.22	0.52	61.26	0.14	362	394	405
No-Ad	0.719	232.13	0.51	53.23	0.11	319	346	354
No-Ad*	0.728	247.58	0.52	57.15	0.12	341	370	379
50-40	0.718	198.50	0.59	46.27	0.13	274	299	307
50-40*	0.730	217.55	0.61	51.29	0.14	302	329	339

Table 1 Characteristics of a best of 5 tiebreak sets match when p_A=0.66 and p_B=0.62

The more relevant observations that can be made from Table 1 include

1. The 50-40 and 50-40* games, whilst producing more efficient match systems than the other game structures, reduce the mean duration by an amount that would appear to be excessive and undesirable *for Grand Slam tennis*.
2. The probability that player A wins the match is slightly reduced (relative to Ad games) when No-Ad* is used and reduced further under the No-Ad system.
3. The No-Ad and No-Ad* games produce similar efficiencies to the present matches using the Ad game. They reduced the means, the standard deviations, the skewness and the CDs.
4. The No-Ad* game produces a best of 5 set scoring system with characteristics roughly midway between those of the Ad game and the No-Ad game. Its mean duration is 13.6 points fewer than the present system, and its CD99.5 is 26 points smaller. It might be considered a useful solution to the issue being studied in this research.

2. Secondly, we consider a men's singles Grand Slam match in which the advantage of serving for both players are less than the Grand Slam average. These parameters could represent a typical match at the French Open between two players with weaker or less successful serves. Table 2 gives the relevant characteristics.

B5 sets	P(A wins)	Mean	Efficiency	Stand Dev	Skew	CD95	CD99	CD99.5
Ad games	0.741	260.68	0.58	61.82	0.15	363	396	407
No-Ad	0.721	229.69	0.55	52.92	0.11	316	343	352
No-Ad*	0.732	245.74	0.57	57.13	0.13	340	368	378
50-40	0.715	196.29	0.60	45.68	0.13	271	295	304
50-40*	0.728	215.15	0.63	50.71	0.14	299	326	335

Table 2 Characteristics of a best of 5 tiebreak sets match when p_A=0.62 and p_B=0.58

Whilst all of the observations made with respect to Table 1 apply also to Table 2, perhaps the most relevant comparison is the observation that P(A wins) decreases for both the 50-40 and the 50-40* games relative to Table 1 (whilst it increases for the other types of games). This is not a surprise as the advantage of serving is reduced with these parameter values.

3. Thirdly, we consider a men's singles Grand Slam match in which the advantage of serving is greater than average. These parameters could represent a typical match at Wimbledon between two players with stronger or more successful serves.

B5 sets	P(A wins)	Mean	Efficiency	Stand Dev	Skew	CD95	CD99	CD99.5
Ad games	0.725	263.30	0.45	60.98	0.12	364	394	404
No-Ad	0.717	236.11	0.46	53.86	0.11	324	350	359
No-Ad*	0.723	251.03	0.46	57.50	0.11	345	373	382
50-40	0.721	201.50	0.57	46.97	0.13	279	303	312
50-40*	0.732	220.49	0.57	51.87	0.14	306	334	343

Table 3 Characteristics of a best of 5 tiebreak sets match when p_A=0.70 and p_B=0.66

Whilst all of the observations made with respect to Table 1 apply also to Table 3, perhaps the most relevant comparison is the observation that P(A wins) increases for both the 50-40 and the 50-40* games relative to Table 1 (whilst it decreases for the other types of games). This is not a surprise as the advantage of serving is enhanced with these parameter values.

Conclusions

The statistical characteristics of five different best of 5 tiebreak sets scoring systems have been studied. The aim of the study was to see whether there was an alternative to the present system using advantage games that might lead to less occurrences of very long matches and thus might be of use in Grand Slam tennis. Several measures for comparing tennis scoring systems have been outlined.

The effect of five different types of games within the best of five sets structure has been analysed. The types of games included the Ad game and the No-Ad game as defined in the Rules of Tennis. The No-Ad* game was also considered. In this game the best of three points is played if deuce is reached. The 50-40 game in which the server needs to win 4 points whilst the receiver needs to win just 3 points in order to win the game, was also considered. The 50-40* game, a modification of the 50-40 game in which the best of 3 points is played if 40-30 is reached, was also considered.

Whilst the 50-40 and 50-40* games were shown to be typically quite efficient for many matches and very effective at reducing match length, they would appear to 'go too far' for consideration at the Grand Slam level. Further, they may appear problematic to some players due to their 'unbalanced' structure.

The No-Ad*, having characteristics somewhere between the Ad game and the No-Ad game resulted in a useful decrease in the number of very long best of 5 tiebreak sets matches. It would appear to be a useful addition to available tennis scoring systems.

References

Barnett T (2012). Analyzing tennis scoring systems: from the origins to today. Journal of Medicine and Science in Tennis 17(2), 68-77.
Barnett T (2016). A recursive approach to modelling the amount of time played in a tennis match. Journal of Medicine and Science in Tennis, 21(2).
Bialik C. (2016). Serving is a disadvantage in some Olympic Sports.
https://fivethirtyeight.com/featurse/serving-is-a-disadvantage-in-some-olympic-sports/

Cross, R. and Pollard, G. H. (2009). Grand Slam men's singles tennis 1991-2009, Service speed and other related data. ITF Coaching and Sports Science Review, 16(49), 8-10.
Cross, R. and Pollard, G. H. (2011). Grand Slam men's singles tennis 1995-2009, Part 2: Points, games and sets. ITF Sports Coaching and Sports Science Review, 53(19), 3-6.
Miles R. E. (1984). Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). J. R. Statist. Soc. B, 46, 93-108.
Pollard G. H. (1983). An analysis of classical and tiebreaker tennis. Australian Journal of Statistics, 25(3), 496-505.
Pollard, G. H. and Barnett, T. (2018). Some new 'short games' within a set of tennis. International Journal of Computer Science in Sport. 17(1), 67-76.
Pollard. G. H. and Noble, K. (2004). The benefits of a new game scoring system in tennis, the 50-40 game. Proceedings of the seventh Australasian conference on mathematics and computers in sport,
Pollard. G. H., Pollard G. N., Lisle, I. and R. Cross. (2010). Bias in Sporting Match Statistics. Proceedings of the Tenth Australasian conference on mathematics and computing in sport, Darwin, edited by A, Bedford and M. Ovens. Mathsport (ANZIAM), 221-228, July, 2010.