Some alternative men's doubles scoring systems

Graham Pollard[1], Tristan Barnett[2], Alan Brown[2] and Geoff Pollard[2]
[1]*University of Canberra*
[2]*Swinburne University of Technology*

In 2006 the scoring system for men's doubles was changed. The purpose of the change was to have matches of shorter and more predictable duration. In this paper the statistical characteristics of the previous and the current scoring systems are quantified. Further, some alternative scoring systems making use of recent ideas are considered, and compared with the current system. The methods used in this paper to produce the results are new, and could be used for a wide range of other sports and other scoring systems.

INTRODUCTION

In 2006 there was a change to the best-of-three sets scoring system used for men's doubles in a range of professional tournaments. The main objectives of the change would appear to have been to reduce somewhat the average length of a match, to play matches that have a more predictable duration, and to reduce the likelihood of 'long' matches. The system now used for these tournaments is a best-of-three sets system with the first two tiebreak sets using no-ad games, and the third set being simply a first-to-ten points match-deciding tiebreak game. Under these changes there is clearly a decrease in the average length of matches, in the variability of the length of matches, and in the likelihood of long matches. Doubles matches using this scoring system have a more predictable duration, and this is beneficial for the players, the spectators and the tournament organizers. With this scoring system in use, a greater proportion of the top men's singles players might play doubles.

In this paper it is assumed that the probability that the better pair wins a point when one of them is serving is p_a, and the probability that the other pair win a point when serving is p_b. For the Australian Open in 2006, the server in men's doubles won 63% of the points. For Wimbledon in 2006, the corresponding percentage was at least 65%.

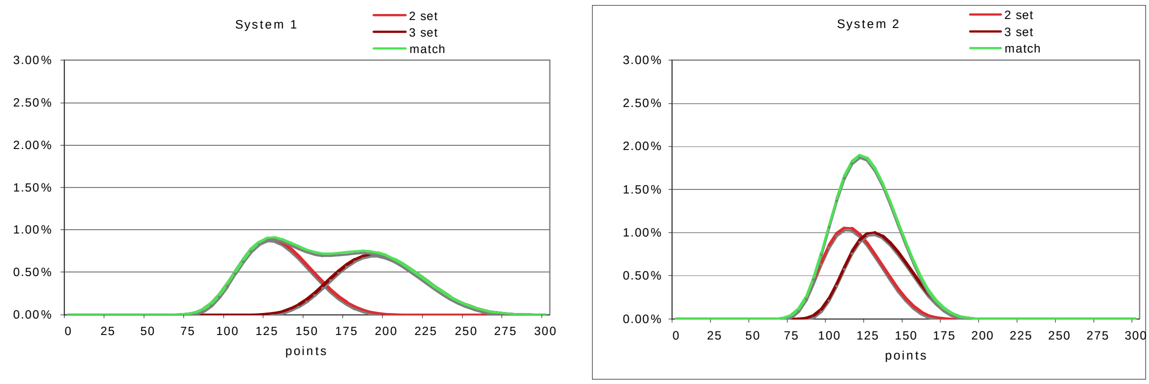

Figure 1. Comparison of the distributions of points in a match for two scoring systems; (a) previous, (b) current; probability of server winning a point, $p_a = p_b = 0.65$.

Several effects of this change to the scoring system are quantified in this paper. Changes to the mean, the variance and the distribution of the number of points in a match are evaluated. Changes to the probability that each pair wins the match are also noted, as are changes to the length of 'long' matches.

In recent times several observations on tennis and on tennis scoring have been reported in the literature. Some of these observations have the potential to be useful in addressing the very issues *that the above change in the men's doubles' scoring addressed. These observations are discussed in the following paragraphs*.

The first observation concerns the usefulness of the '50-40 game', particularly in men's doubles (see Pollard and Noble, 2004). Note that in a 50-40 game, the server is required to reach 50 (one more point than 40) before the receiver reaches 40, in order to win the game, whilst the receiver wins the game by reaching 40. Pollard and Noble (ibid) noted that, for p-values relevant to men's doubles (values typically greater than 0.6), 'the longest best-of-three tiebreak sets matches (as measured by the 98% point in the distribution of the number of points played) can be reduced by about 30 points by using 50-40 games instead of no-ad games'. They also noted that for such p-values 'the probability player A wins a best-of-three tiebreak sets match using 50-40 games is comparable to when using no-ad games, even though about 20 less points are required on average'. Note that the seventh point in the no-ad game creates a lack of symmetry with respect to the first and second court and a potential for unfairness, and that these two characteristics are removed by using the 50-40 game. Thus, it would appear that the '50-40' game has some merit relative to the no-ad game.

A second observation is that first-to-six games sets can be unfair in doubles (Pollard, 2005). This unfairness results from the fact that one player in a pair may have three service games in a set whilst his partner may have just two. This unfairness can be removed by playing (say) 'first-to-seven games' sets, which operate probabilistically like 'best-of-twelve games' sets in which each player has 3 serves each (alternatively, the unfairness can be removed by playing 'first-to-five games' sets, which operate probabilistically like 'best-of-eight games' sets in which each player has 2 serves each). These ('longer') 'first-to-seven games' sets could work well when used in conjunction with the shorter '50-40' games, and so they are considered in this paper. Further, it can be seen that 'first-to-five games' sets might work well with '60-50' games (a natural extension of 50-40 games).

Thirdly, the present first-to-seven points (leading by at least two points) and first-to-ten points (leading by at least two points) tiebreak games can also be unfair in doubles (Pollard (2005)). (For example, the first twelve points in a 'first-to-seven' points tiebreak game works probabilistically like a best-of-twelve points structure with the player who serves second in the tiebreak game having 4 serves and his partner having only 2 serves, whilst the other two players have 3 serves each. This can lead to unfairness.) He showed that this unfairness can be removed by playing first-to-nine points, leading by at least 2 points, and if 8-8 is reached, first-to-thirteen points leading by at least 2 points, and if 12-12 is reached, first-to-seventeen points leading by at least 2 points,…. If the shorter 50-40 games were used, it might be practical to use this fairer and somewhat longer 'first-to-9 points, …' tiebreak game as well.

Fourthly, Pollard and Noble (2003) considered the idea of declaring a set a draw if the games' score reaches 6-6. Accordingly, the set score is incremented by 2, 1 or 0 for a win, draw or loss respectively. They considered 'best-of-two sets' matches with the winner of the match being the first player to reach a set score at least 3. If a set score of 2-2 was reached, a (long) tiebreak game was played to determine the winner. Interestingly, they noted that 'when playing no-ad games and best-of-two sets with a TB(7,2) deciding match tiebreak, it is (perhaps surprisingly) more efficient not to play the tiebreak game at 6-6 than it is to play it'. We note here that declaring a set to be a draw at 6-6, although fair for the players, can be seen as an unattractive rule for those spectators who like the excitement of the tiebreak game. Then again, for most non-professional matches (and even for some professional matches), there are no spectators. Thus, such a scoring system can have its uses.

Fifthly, Pollard and Barnett (2006) considered situations where psychological factors might exist, and concluded that 'for the situation in which players have a psychological advantage when ahead in games' score, the player who serves first in a set of tennis has…an advantage. This (set) advantage can be decreased by alternating service at the beginning of each set (with the exception of the final third or fifth set which would be determined under the present rules'. It would seem reasonable to apply this 'alternating-service rule across sets' in all tennis matches, not just doubles matches. These authors also suggested a minor adjustment to the service order in the third set of a best-of-three sets doubles match (ibid, page 196).

Sixthly, Pollard and Barnett (ibid) concluded that (in the presence of psychological effects) 'the advantage of serving first in a set can be further reduced by allowing the player who serves in the second game of the set to serve on two consecutive games at some stage within that set'. This observation applies to doubles as well as singles. Its use could be quite exciting for the spectators, as well as adding strategic elements for the players.

Finally, Pollard and Noble (2006) considered 13 elementary countback rules and concluded that 'countback methods at 6-6 in a set of tennis can be more effective at correctly identifying the better player than is the playing of a tiebreak game. This is particularly so when serving is a strong advantage in a match, and this occurs in many doubles…situations.' They also concluded that 'it would seem that the most relevant occasions on which a countback might be used are when there are no spectators.' They also considered very simple countback rules that they called auxiliary scoring systems, and conclude that the better auxiliary scoring systems 'perform quite well in the situations where they might be considered'. This countback and auxiliary scoring systems approach to allocating a set outcome at 6-6 is not considered further in this paper, but the reader is referred to the above paper for a quantitative assessment of the effectiveness of countback systems.

The fifth and sixth observations can be included in any modifications, if thought appropriate, without doing any further mathematics. Also, the final observation could be incorporated into the work of this paper by plugging in the results of Pollard and Noble (2006) into the work being reported in this paper.

Thus, putting these seven observations together, it would seem reasonable to consider a few best-of-three sets scoring systems making use of some or all of the first four observations, namely, '50-40' games, 'first-to-seven games' tiebreak sets for the first two sets, 'first-to-nine, …' points tiebreak games, and drawn sets at 6-6. The characteristics of some scoring systems using these constructs could be compared with those of the new system and the previous system.

What are the desirable (statistical) characteristics of a good tennis scoring system? The 'three-nesting' aspect of tennis (points, games and sets) is taken as a given or fixed part of tennis scoring systems, as it allows either player to overcome a period of poor play. Games and sets might be made longer or shorter than at present if there is an advantage in doing so. It is fundamental that a scoring system should have an appropriate average number of points played, and an appropriate value for the probability that the better player wins. Also, the standard deviation of the number of points played should not be too large, so that matches have a reasonably predictable duration. Strongly positively skewed distributions of duration are to be avoided as, under such scoring systems, very long matches can result, and this can delay the matches that follow, and can lead to unfairness in the tournament setting. Scoring systems with good efficiency at correctly identifying the better player are preferable to ones with not-so-good efficiency. Given these four or five characteristics to be considered before adopting an appropriate scoring system for a tournament, there are usually compromises that need to be made between them in choosing one particular scoring system over another.

METHODS

The probability that the better player wins the match, and the mean and higher moments of the number of points in a match, were evaluated exactly using recursion methods on a spreadsheet. Details of these recursion methods are omitted here, but will be available (Brown, Barnett and Pollard, 2008).

The estimation of the distribution of the number of points in a match requires some care since, as seen in Figure 1, it may be bi-modal, depending upon the design of the third set. Furthermore a distribution is not uniquely determined by its moments. To overcome these difficulties, the distributions for matches requiring 2 sets or 3 sets to complete were estimated separately. This was done using the first four moments of each distribution and the Normal Power approximation (Pesonen, 1975). The Normal Power approximation requires a basic assumption that the distribution is uni-modal, and it would be inappropriate to use it in conjunction with the statistics for an overall match reported in the tables below. The probability-weighted sum of the two distributions was used to estimate the distribution in the overall match, from which estimates of the 98% and 2% points were obtained by interpolation. These two points in the distributions are used as measures of 'long' and 'short' matches for the various scoring systems. The results were checked (and agreed) with the best-of-three tiebreak sets results in the paper by Pollard and Noble (2004), who used simulations of 1,000,000 matches.

We note here that two scoring systems can be compared for their efficiencies at correctly identifying the better pair. Thus, given two scoring systems with the same expected number of points played in a match, scoring system 1 is said to be more efficient than scoring system 2 if scoring system 1 has a higher value for the probability that the better player wins the match. The efficiencies of scoring systems with differing values for the expected number of points played can also be evaluated (Miles, 1984). It is noted here that scoring systems with high efficiencies (that is, efficiencies close to 1) typically have particularly large standard deviations, and hence are not appropriate in the sporting context.

Characteristics of the following seven best-of-three sets scoring systems were evaluated for values of p_a and p_b near 0.6, 0.65 and 0.7. These values should cover most of the matches played by professional men.

1. (The previous system) The standard best-of-three tiebreak sets using advantage games.
2. (The current system) The first two sets are tiebreak sets, use no-ad games (only one point is played at deuce), and use the first-to-seven points (leading by at least two points) tiebreak games if 6-6 is reached. The third set is simply a match deciding first-to-ten points (leading by at least two points) tiebreak game.
3. The third system considered is the current system (system 2. above), modified in two ways, namely 50-40 games are used instead of no-ad games, and 'first-to-seven games' sets are used instead of standard tiebreak sets.
4. The fourth system considered is the third system above modified in one way, namely that all tiebreak games played are 'first-to-nine points, …' tiebreak games.
5. The fifth system is just the fourth system modified in two ways. The first two sets are 'first-to-five games' sets, and games are '60-50' games. Thus, this system has 'longer' games, but 'shorter' sets.
6. The sixth system uses a 'first-to-seven' games structure of a set (as in system 3.), but it involves a new structure within a game that allows its outcome to be a win, a loss or a draw. With 'draw-game', player A1 serves 6 points, ends are changed, and player B1 serves (up to) 6 points. The first pair to win 7 points in the 'draw-game' wins the game, and increments its game score by 2. If the points' score reaches 6-6, the game concludes as a draw and each pair increments its game score by 1. The second 'draw-game'

consists of player A2 serving 6 points, change of ends, followed by player B2 serving (up to) 6 points. (It can be seen that occasionally near the end of a set the 'draw-game' being played may not need to be finished as one pair may won enough points to reach a games' score of 7 and claim the set. (Our method of calculation addresses this issue of early termination of a 'draw-game'.) Pair B should start the serving in the second set (B2 followed by A2 would be better than starting with B1 followed by A1 or A2).

7. The seventh system considered is the third system above modified in a way that allows the outcome of a set to be a win, loss or draw. The first two sets are played as a 'draw-set' where the first pair to win 7 games wins the set, and increments its set score by 2. If the games' score reaches 6-6, we have a draw and each pair increments its set score by 1. The winner of the match is the first pair to reach a sets' score of 3. If the sets' score reaches 2-2, the 'first-to-nine points, … ' tiebreak game is played. (It is noted that for this system the match can sometimes be over before the second set is completed. Our method of calculation addresses this issue of early termination of a 'draw-set'.)

A comparison of the current and previous scoring systems

p_a	0.60	0.62	0.64	0.65	0.67	0.69	0.70	0.72	0.74
p_b	0.60	0.58	0.56	0.65	0.63	0.61	0.70	0.68	0.66
P	0.5000	0.6974	0.8491	0.5000	0.6893	0.8380	0.5000	0.6795	0.8243
μ	164.0	160.0	149.6	164.0	160.4	151.0	165.5	162.4	154.1
σ	41.2	41.4	40.8	40.7	40.8	40.3	40.4	40.6	40.4
sk	0.27	0.34	0.55	0.26	0.33	0.52	0.22	0.29	0.47
ku	-0.71	-0.69	-0.49	-0.77	-0.74	-0.56	-0.87	-0.85	-0.70
ρ	N/A	0.6179	0.6030	N/A	0.5343	0.5213	N/A	0.4355	0.4251
98%	248.0	245.6	238.9	246.1	244.2	238.5	245.2	243.8	239.5
2%	95.0	93.3	87.5	96.2	94.5	90.0	98.3	96.0	92.7

Table 1. Characteristics of System 1 (the previous system).

p_a	0.60	0.62	0.64	0.65	0.67	0.69	0.70	0.72	0.74
p_b	0.60	0.58	0.56	0.65	0.63	0.61	0.70	0.68	0.66
P	0.5000	0.6583	0.7923	0.5000	0.6579	0.7916	0.5000	0.6577	0.7915
μ	123.4	122.0	118.0	125.0	123.6	119.8	127.9	126.6	122.9
σ	20.3	20.5	20.8	20.3	20.5	20.7	20.4	20.6	20.9
sk	0.35	0.35	0.39	0.34	0.34	0.39	0.26	0.27	0.32
ku	-0.16	-0.16	-0.15	-0.25	-0.24	-0.21	-0.40	-0.40	-0.37
ρ	N/A	0.5103	0.4961	N/A	0.4745	0.4605	N/A	0.4268	0.4133
98%	167.3	166.2	163.3	168.9	167.9	164.8	171.7	170.8	168.2
2%	85.6	85.2	81.9	87.8	85.6	84.5	91.0	88.6	85.2

Table 2. Characteristics of System 2 (the current system).

Tables 1 and 2 give some characteristics of the previous scoring system (system 1) and the current scoring system (system 2) for values of (p_a, p_b) equal to (0.60,0.60), (0.62,0.58), (0.64,0.56), …, (0.74,0.66). P = probability better pair win. μ, σ, sk and ku denote mean, standard deviation, skewness and kurtosis, respectively, of the number of points in a match. ρ = efficiency. 98% and 2% denote percentiles of the distribution. It can be seen from the tables, and figure 1 above, that

(i) the expected number of points played for the current system is about 35 points less than for the previous system.

(ii) the standard deviation of the number of points in a match for the current scoring system is approximately half that of the previous system. This is a substantial reduction.

(iii) using the 98% points (i.e. the number of points such that 2% of matches played involve a greater number of points) of the distributions as a measure of ‘long’ matches for each scoring system, ‘long’ matches are approximately 75 points shorter for the current system. This is a substantial reduction achieved by using the current system.

(iv) the current system is less efficient than the previous system, particularly for values of p_a and p_b near 0.6.

The probability that the better pair win the match is reduced by up to almost 0.06 (in the tables) by using the current system. This is the price paid by using a less efficient system of shorter duration. It can be seen that that the current scoring system has a considerably shorter expected duration than that of the previous one, and its standard deviation of duration is substantially smaller. ‘Long’ matches under the current scoring system are very much shorter than ‘long’ matches under the previous system. Thus, under the current system, matches are certainly of a more predictable duration. The downside of the current system is that it is somewhat less efficient, and that there is a reduction in the likelihood of the better player winning. Overall however, the change would appear to have met its objectives.

A comparison of systems 3 to 7 with the current system (system 2)

Figure 2 together with tables 3, 4 and 5 give relevant results for comparing systems 3, 4, 5, 6 and 7 with system 2.

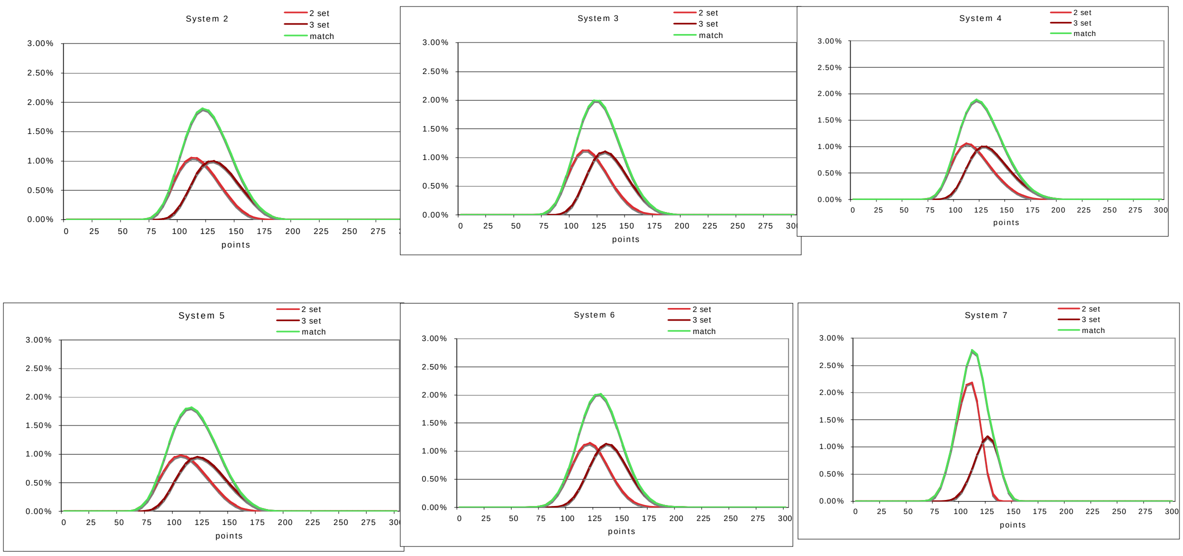

Figure 2. Comparison of the distributions of points in a match for six scoring systems; (a) current, (b) … (f) new; probability of server winning a point, $p_a = p_b = 0.65$.

	P	M	σ	sk	ku	ρ	98%	2%
1	0.6974	160.0	41.4	0.34	-0.69	0.6179	245.6	93.3
2	0.6583	122.0	20.5	0.35	-0.16	0.5103	166.2	85.2
3	0.6626	122.2	19.6	0.32	-0.08	0.5385	164.5	85.5
4	0.6640	122.9	21.2	0,54	0.29	0.5450	169.9	85.1
5	0.6583	115.6	22.3	0.44	-0.11	0.5386	163.9	75.0
6	0.6675	127.8	19.7	0.18	-0.12	0.5478	169.9	91.5
7	0.6591	109.1	14.8	-0.05	-0.29	0.5762	139.9	82.5

Table 3. Characteristics of the scoring systems when $p_a = 0.62$ and $p_b = 0.58$.

	P	M	σ	sk	ku	ρ	98%	2%
1	0.6893	160.4	40.8	0.33	-0.74	0.5343	244.2	94.5
2	0.6579	123.6	20.5	0.34	-0.24	0.4745	167.9	85.6
3	0.6660	123.8	19.8	0.34	-0.08	0.5259	166.8	87.4
4	0.6676	124.7	21.5	0.56	0.29	0.5328	172.4	86.8
5	0.6605	118.0	22.7	0.43	-0.18	0.5143	167.1	76.7
6	0.6722	127.9	19.8	0.17	-0.10	0.5497	170.4	91.0
7	0.6620	110.2	14.5	-0.05	-0.27	0.5617	140.6	85.4

Table 4. Characteristics of the scoring systems when $p_a = 0.67$ and $p_b = 0.63$.

	P	M	σ	sk	ku	ρ	98%	2%
1	0.6795	162.4	40.6	0.29	-0.85	0.4355	243.8	96.0
2	0.6577	126.6	20.6	0.27	-0.40	0.4268	170.8	88.6
3	0.6699	126.1	19.9	0.37	-0.09	0.5002	169.6	90.7
4	0.6719	127.2	22.0	0.58	0.25	0.5080	175.7	89.6
5	0.6628	122.0	23.3	0.37	-0.34	0.4731	172.2	80.0
6	0.6795	128.0	20.1	0.17	-0.06	0.5524	171.1	90.4
7	0.6647	111.2	13.9	-0.03	-0.22	0.5314	140.7	87.1

Table 5. Characteristics of the scoring systems when $p_a = 0.72$ and $p_b = 0.68$.

It can be seen from these figures and tables that

(i) System 3 has similar values to system 2 for the mean, standard deviation, skewness, the 98% point, and the 2% point. System 3 has somewhat larger values for the probability that the better player wins, and for the efficiency. Thus, system 3 would appear to slightly better than system 2 with respect to the desirable characteristics of a tennis scoring system.

(ii) System 4 has similar values to system 2 for mean and standard deviation, whilst it has a larger value for skewness and a slightly larger 98% point. System 4 is more efficient than system 2. For example, in Table 5, system 4 is $(0.5080/0.4268 - 1) \times 100$ percent or 19.0% more efficient than system 2. System 4 has a higher P value than system 2.

(iii) System 5 has a slightly smaller mean, and a slightly larger standard deviation and skewness. System 5 has a similar 98% point, but a reduced 2% point indicating that 'short' matches are even somewhat shorter. This is probably not a good thing in terms of players being able to recover from a period of poor play. System 5 has a slightly higher P value, and greater efficiency.

(iv) System 6 has a slightly larger mean, a similar standard deviation, a slightly smaller skewness, and a similar value for the 98% point. System 6 has greater efficiency and a higher value of P.

(v) System 7 is somewhat different to the others. It has a smaller mean and standard deviation, a considerably smaller skewness, and its 98% point is a lot less than that of system 2. It has greater efficiency than system 2 and a larger value for P.

RESULTS AND DISCUSSION

All of the five alternative scoring systems would appear to be just as good as, or even a little better than the current scoring system from a statistical point of view. They are all at least slightly more efficient, and have higher values for P, the probability that the better player wins. Overall, there is not really a great deal of difference between these various systems, as they all use a modified third set. System 6 using 'draw-games' has one of the highest efficiencies, but its game structure is quite different to that of other systems. System 7 using ‘draw-sets’ is perhaps the most interesting one. It has a reduced mean, a small standard deviation, a low 98% point, and good efficiency with an improved value of P.

CONCLUSIONS

The objectives of adopting a new scoring system for men’s doubles in 2006 would appear to have been to reduce somewhat the average length of matches, to have matches of a more predictable duration, and to reduce the likelihood of long matches. These objectives have been met. However, a slight downside of the current scoring system is that it is somewhat less efficient, with a smaller value for the probability that the better player wins.

In this paper five alternative scoring systems making use of some recent ideas in the literature are considered. Overall, these five scoring systems are quite similar to the current system with respect to the average length of matches, the improved predictability of the duration of matches, and the likelihood of long matches. Each of the five systems however is more efficient than the current system, and has a higher value for the probability that the better player wins. Thus, on statistical grounds, they would appear to be legitimate alternatives to the current system. Of course, grounds other than statistical ones are considered when adopting a new scoring system. Nevertheless, some of the five alternative scoring systems described in this paper may be considered useful additions to tennis scoring systems in general.

ACKNOWLEDGEMENTS

The authors would like to thank KAN-soft for providing data containing the percentage of points won on serve in men’s doubles.

REFERENCES

Brown, A., Barnett, T. and Pollard, G. (2008). A recursion method for evaluating the moments of a nested scoring system. To be reviewed for *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport.* Queensland, Australia: MathSport.

Miles, R.E. (1984).Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society*, **46**(1), 93-108.

Pesonen, E. (1975) NP-technique as a tool in decision making. Astin Bulletin **8**(3), 359-363.

Pollard, G. (2005). On solving an aspect of unfairness in the tennis doubles scoring system. *Chance*, **18**(4), 17-19.

Pollard, G. and Barnett, T. (2006). Fairer service exchange mechanisms for tennis when some psychological factors exist. In *Proceedings of the Eighth Australasian Conference on Mathematics and Computers in Sport* (edited by J. Hammond and N. de Mestre), pp 189-198. Queensland, Australia: MathSport.

Pollard, G. and Noble. K. (2003). Scoring to remove long matches, increase tournament fairness and reduce injuries. *Medicine and Science in Tennis*, **8**(3), 12-13.

Pollard, G. and Noble, K. (2004). The benefits of a new game scoring system in tennis: the 50-40 game. In *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport* (edited by R. Morton and S. Ganesalingam), pp 262-265. New Zealand: Massey University.

Pollard, G. and Noble, K. (2006). Countback methods and auxiliary scoring systems in tennis. In *Proceedings of the Eighth Australasian Conference on Mathematics and Computers in Sport* (edited by J. Hammond and N. de Mestre), pp 170-177. Queensland, Australia: MathSport.