REDUCING THE LIKELIHOOD OF LONG TENNIS MATCHES

Barnett, T.[1], Brown, A.[1] and Pollard, G.[2]
[1] Swinburne University, Melbourne, VIC, Australia
[2] University of Canberra, Canberra, ACT, Australia

ABSTRACT

Long matches can cause problems for tournaments. For example, the starting times of subsequent matches can be substantially delayed causing inconvenience to players, spectators, officials and television scheduling. They can even be seen as unfair in the tournament setting when the winner of a very long match, who may have negative aftereffects from such a match, plays the winner of an average or shorter length match in the next round. Long matches can also lead to injuries to the participating players.

One factor that can lead to long matches is the use of the advantage set as the fifth set, as in the Australian Open, the French Open and Wimbledon. Another factor is long rallies and a greater than average number of points per game. This tends to occur more frequently on the slower surfaces such as at the French Open. The mathematical method of generating functions is used to show that the likelihood of long matches can be substantially reduced by using the tiebreak game in the fifth set, or more effectively by using a new type of game, the 50-40 game, throughout the match.

KEY WORDS

tennis, scoring systems, sport, generating functions, long tennis matches

1. INTRODUCTION

In recent years there have been a number of grand slam matches decided in long fifth sets. In the third round of the 2000 Wimbledon mens singles, Philippoussis defeated Schalken 20-18 in the fifth set. Ivanisevic defeated Krajicek 15-13 in the semi-finals of Wimbledon in 1998. In the quarter-finals of the 2003 Australian Open mens singles, Andy Roddick defeated Younes El Aynaoui 21-19 in the fifth set, a match taking 83 games to complete and lasting a total duration of 5 hours. The night session containing this long match required the following match to start at 1 am. Long matches require rescheduling of following matches, and also create scheduling problems for media broadcasters. They arise because of the advantage set, which gives more chance of winning to the better player (Pollard and Noble, 2002), but has no upper bound on the number of games played. It may be in the interests of broadcasters and tournament organizers to decrease the likelihood of long tennis matches occurring.

Pollard (1983) calculated the mean and variance of the duration of a best-of-three sets match of classical and tiebreaker tennis by using the probability generating function. It is well established that the mean and standard deviation completely describe the normal distribution. When a distribution is not symmetrical about the mean, the coefficients of skewness and kurtosis, as defined in Stuart and Ord (1987), are important to graphically interpret the shape of the distribution. This commonly has been done by using the probability or moment generating function. The cumulant generating function (taking the natural logarithm of the moment generating function), can also be used to calculate the parameters of the distribution in a tennis match. The cumulant generating function is particularly useful for calculating the parameters of distributions for the number of points in a tiebreaker match, since the critical property of cumulant generating functions is that they are additive for linear combinations of independent random variables. The layout of this paper is as follows. For convenience of the less mathematically inclined we defer the presentation of the mathematics of generating functions applied to tennis till Section 3. Instead we will begin in Section 2 with a discussion on several aspects of long matches, relying on graphical results to advance our arguments as to how they might be curtailed. We aim to show that the likelihood of long matches can be substantially reduced by using the tiebreak game in the fifth set, or more effectively by the use of a new type of game, the 50-40 game (Pollard and Noble, 2004), throughout the match. In Section 4 we make some concluding remarks.

2. DISCUSSION OF THE PROBLEM (using graphical results)

Up until 1970 (approx), all tennis sets were played as advantage sets, where to win a set a player must reach at least 6 games and be ahead by at least 2 games. The tiebreaker game was introduced to shorten the length of matches. A tiebreaker game is played when the set score reaches 6-games all. However in three of the four grand slams (Australian Open, French Open and Wimbledon), an advantage set is still played in the deciding fifth set. Figure 1 represents a comparison of a match with 5 advantage sets (5adv), 5 tiebreaker sets (5tie) and 4 tiebreaker sets with a deciding advantage set (4tie1adv). The probability of each player winning a point on serve is given as 0.6 to represent averages in men's tennis. The long tail given by the match with 5adv gives an indication as to why the tiebreaker game was introduced to the tennis scoring system. It is well known that the dominance of serve in men's tennis has increased since the introduction of the tiebreaker game. This creates a problem when two big servers meet in a grand slam event where the deciding fifth set is played as an advantage set. Figure 2 represents a match with 4tie1adv for different values of players winning points on serve. It shows that for two strong servers winning 0.7 of points on serve, there is a long tail in the number of points played. In comparison with Figure 3, which represents a match with 5tie, the tail is substantially reduced for two players winning 0.7 of points on serve. Figure 4 represents a match with 5 tiebreaker sets, where a standard 'deuce' game is replaced by a 50-40 game. It shows an even greater improvement to reducing the number of points played in a match compared to Figure 3. In the 50-40 game the server has to win the standard 4 points, while the receiver only has to win 3 points. Such a game requires at most 6 points.

Figure 1: Distribution of a match with different scoring systems

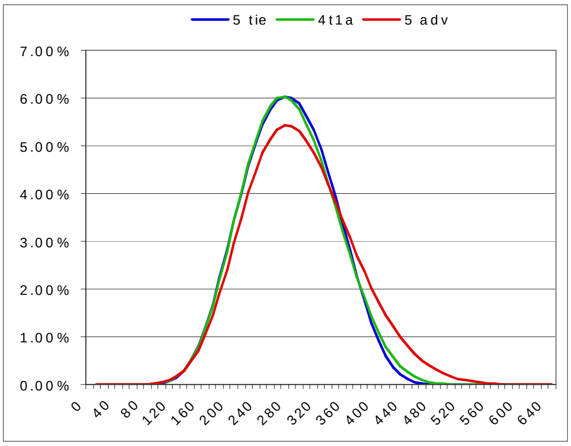


Figure 2: Distribution of an advantage match (4tie1adv) for different values of players winning points on serve

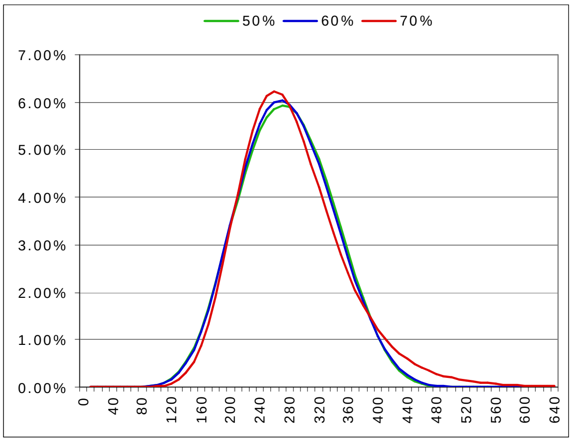


Figure 3: Distribution of a tiebreaker match (5tie) for different values of players winning points on serve

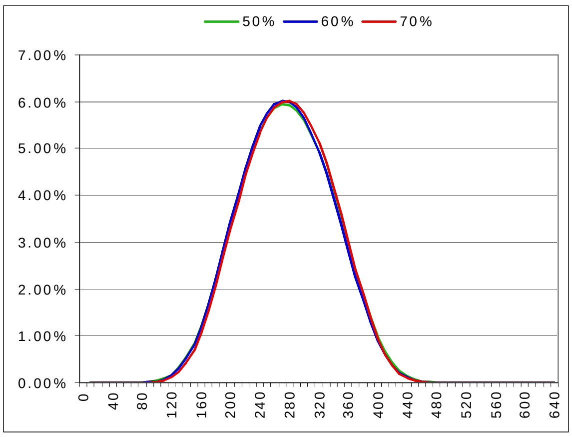
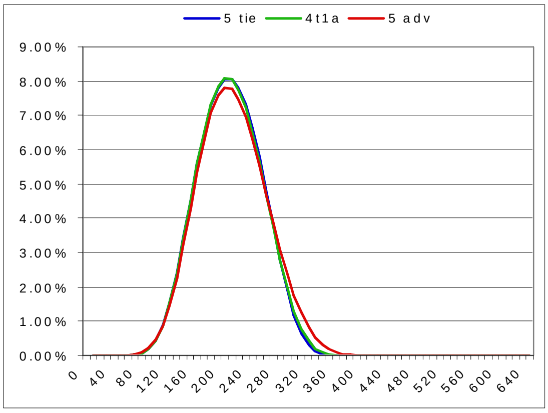


Figure 4: Distribution of a tiebreaker match (5tie) for different values of players winning points on serve, by using 50-40 games instead of standard 'deuce' games



3. THE MATHEMATICS OF GENERATING FUNCTIONS

3.1 MODELLING A TENNIS MATCH

3.1.1 FORWARD RECURSION

The state of a tennis match between two players is represented by a scoreboard. The scoreboard shows the points, games and sets won by each player, and is updated after each point has been played. It is assumed that the conditional probability of the server winning the point depends only on the data shown on the scoreboard. This enables the progress of the match to be modelled using forward recursion. An additional assumption is that the probabilities of each player winning a point on his own service remain constant throughout the match.

3.1.2 DEVELOPMENT OF GENERATING FUNCTIONS OF DISTRIBUTIONS

The forward recursion enables the probabilities of various possible scoreboards to be calculated. These probabilities can be collected in the form of probability generating functions, or moment generating functions (using the transformation $v = e^u$).

Lemma: If X and Y are independent random variables and $Z = X + Y$ then:

$$m_Z(t) = m_X(t) * m_Y(t).$$

It becomes convenient at times to take logarithms, and work in terms of cumulant generating functions, since $K_Z(t) = K_X(t) + K_Y(t)$.

The higher order cumulants depend on powers of the scale for the random variable, and for the purposes of communication it is useful to transform them into non-dimensional statistics (i.e. numbers) such as the coefficients of variation, skewness and kurtosis.

3.1.3 THE INVERSION OF THE CUMULANTS USING NORMAL POWER APPROXIMATION

This gives a continuous approximation to a discrete distribution (Pesonen, 1975). The formula is asymptotic and works reasonably well for unimodal distributions with the coeffiecient of skewness less than 2 and the coefficient of kurtosis less than 6. i.e. tails die off at least as fast as the exponential distribution.

3.2 THE NUMBER OF POINTS IN A GAME

Let X be a random variable of the number of points played in a game. Let $f^{pg}{}_A\ (x)$ represent the distribution of the number of points played in a game for player A serving, where $f^{pg}{}_A\ (x) = P(X = x)$. This gives the following:

$f^{pg}{}_A\ (4) = N^{pg}{}_A\ (4,\ 0) + N^{pg}{}_A\ (0,\ 4)$
$f^{pg}{}_A\ (5) = N^{pg}{}_A\ (4,\ 1) + N^{pg}{}_A\ (1,\ 4)$
$f^{pg}{}_A\ (6) = N^{pg}{}_A\ (4,2) + N^{pg}{}_A\ (2,\ 4)$
$f^{pg}{}_A\ (\mathrm{x}) = N^{pg}{}_A\ (3,\ 3)[p^2{}_A + (1 - p_A)^2][2p_A(1 - p_A)]^{(x-8)/2}$, if $x = 8,\ 10,\ 12,\ .....$

where:
$N^{pg}{}_A(a,b)$ represents the probability of reaching point score (a,b) in a game for player A serving.
p_A represents the probability of player A winning a point on serve.

Croucher (1986) gives algebraic expressions for calculating $N^{pg}{}_A\ (a,b)$.

Let $m(t)$ denote the moment generating function X. Generating functions can be used to describe a distribution, such as $f^{pg}{}_A(x)$ for all x. It is well established (Stuart and Ord,1987) that the mean, variance, coefficient of skewness and coefficient of kurtosis of X can be obtained from generating functions.

The moment generating function for the number of points in a game for player A serving, $m^{pg}{}_A\ (t)$, becomes:

$$\sum_x e^{tx} f^{pg}{}_A(x) = e^{4t} f^{pg}{}_A(4) + e^{5t} f^{pg}{}_A(5) + e^{6t} f^{pg}{}_A(6) + [N^{pg}{}_A(3,3)(1 - N^{pg}{}_A(1,1))e^{8t}] \ /\ [1 - N^{pg}{}_A(1,1)e^{2t}]$$

The mean number of points in a game $M^{pg}{}_A$, with the associated variance $V^{pg}{}_A$ are calculated from the moment generating function using *Mathematica* and given as:

$$M^{pg}{}_A = \frac{4\{p_A(1-p_A)[6p_A^2(1-p_A)^2 - 1] - 1\}}{[1 - 2p_A(1-p_A)]}$$

$$V^{pg}{}_A = \frac{4p_A(1-p_A)[1 - p_A(1-p_A)(1 - 12p_A(1-p_A)(3 - p_A(1-p_A)(5 + 12p_A^2(1-p_A)^2)))]}{[1 - 2p_A(1-p_A)]^2}$$

Similar expressions can be obtained for the coefficient of skewness $S^{pg}{}_A$, and the coefficient of kurtosis $K^{pg}{}_A$.

Let $U^{pg}{}_A$ represent the standard deviation of the number of points in a game for player A serving. Let $C^{pg}{}_A$ represent the coefficient of variation of the number of points in a game for player A serving. It follows that $U^{pg}{}_A = \sqrt{V^{pg}{}_A}$ and $C^{pg}{}_A = U^{pg}{}_A / M^{pg}{}_A$.

Table 1: The parameters of the distributions of points in a game for different values of p_A

p_A	$M^{pg}{}_A$	$U^{pg}{}_A$	$C^{pg}{}_A$	$S^{pg}{}_A$	$K^{pg}{}_A$
0.50	6.75	2.77	0.41	2.16	6.95
0.55	6.68	2.73	0.41	2.17	7.01
0.60	6.48	2.59	0.40	2.20	7.21
0.65	6.19	2.37	0.38	2.25	7.59
0.70	5.83	2.10	0.36	2.34	8.25
0.75	5.45	1.78	0.33	2.46	9.27

Table 1 represents $M^{pg}{}_A$, $U^{pg}{}_A$, $C^{pg}{}_A$, $S^{pg}{}_A$ and $K^{pg}{}_A$ for different values of p_A. Notice that the mean and standard deviation are greatest when $p_A = 0.50$, but the coefficients of skewness and kurtosis are greatest when p_A approaches 1 or 0. The generating functions to follow are for player A serving first in the tiebreaker game or set.

The moment generating function for the number of points in a tiebreaker game, $m^{pgT}{}_A(t)$ becomes:

$$m^{pgT}{}_A(t) = e^{7t}f^{pgT}{}_A(7)+e^{8t}f^{pgT}{}_A(8)+e^{9t}f^{pgT}{}_A(9)+e^{10t}f^{pgT}{}_A(10)+e^{11t}f^{pgT}{}_A(11)+e^{12t}f^{pgT}{}_A(12)+ N^{pgT}{}_A(6,6)(1-N^{pgT}{}_A(1,1))e^{14t}/[1-N^{pgT}{}_A(1,1)e^{2t}]$$

where:
$f^{pgT}{}_A(x)$ represents the distribution of the number of points played in a tiebreaker game.
$N^{pgT}{}_A(a,b)$ represents the probability of reaching point score (a,b) in a tiebreaker game.

The moment generating functions for the number of games in a tiebreaker set, $m^{gsT}{}_A(t)$ and advantage set, $m^{gs}{}_A(t)$ become:

$$m^{gsT}{}_A(t)=e^{6t}f^{gsT}{}_A(6)+e^{7t}f^{gsT}{}_A(7)+e^{8t}f^{gsT}{}_A(8)+e^{9t}f^{gsT}{}_A(9)+e^{10t}f^{gsT}{}_A(10)+e^{12t}f^{gsT}{}_A(12)+ e^{13t}f^{gsT}{}_A(13)$$

$$m^{gs}{}_A(t) = e^{6t}f^{gs}{}_A(6)+e^{7t}f^{gs}{}_A(7)+e^{8t}f^{gs}{}_A(8)+e^{9t}f^{gs}{}_A(9)+e^{10t}f^{gs}{}_A(10)+ N^{gs}{}_A(5,5)(1-N^{gs}{}_A(1,1))e^{12t}/[1-N^{gs}{}_A(1,1)e^{2t}]$$

where:
$f^{gsT}{}_A(x)$ represents the distribution of the number of games played in a tiebreaker set.
$f^{gs}{}_A(x)$ represents the distribution of the number of games played in an advantage set.
$N^{gs}{}_A(c, d)$ represents the probability of reaching (c,d) in an advantage set.

3.3 THE NUMBER OF POINTS IN A SET

3.3.1 THE PARAMETERS OF DISTRIBUTIONS OF THE NUMBER OF POINTS IN A SET

Let $m^{pg}{}_{A+}(t)$ and $m^{pg}{}_{A-}(t)$ be the moment generating functions of the number of points in a game when player A wins and loses a game on serve respectively. Let $m^{pg}{}_{B+}(t)$ and $m^{pg}{}_{B-}(t)$ be the moment generating functions of the number of points in a game when player B wins and loses a game on serve respectively. Let $s(c, d)$ be the moment generating function of the number of points in a set conditioned on reaching game score (c,d). It can be shown that
$s(6,1) = 3[m^{pg}{}_{A+}(t)]^3[m^{pg}{}_{B-}(t)]^2[m^{pg}{}_{A+}(t)m^{pg}{}_{B+}(t) + m^{pg}{}_{A-}(t)m^{pg}{}_{B-}(t)]$ and
$s(1,6) = 3[m^{pg}{}_{A-}(t)]^3[m^{pg}{}_{B+}(t)]^2[m^{pg}{}_{A+}(t)m^{pg}{}_{B+}(t)+m^{pg}{}_{A-}(t)m^{pg}{}_{B-}(t)]$.
Similar conditional moment generating functions can be obtained for reaching all score lines (c, d) in a set. The moment generating function for the number of points in a tiebreaker set becomes:

$$m^{psT}{}_{A}(t) = N^{gsT}{}_{A}(6,0)s(6,0) + N^{gsT}{}_{A}(6,1)s(6,1) + N^{gsT}{}_{A}(6,2)s(6,2) + N^{gsT}{}_{A}(6,3)s(6,3) + N^{gsT}{}_{A}(6,4)s(6,4)+N^{gsT}{}_{A}(7,5)s(7,5)+N^{gsT}{}_{A}(0,6)s(0,6)+N^{gsT}{}_{A}(1,6)s(1,6)+N^{gsT}{}_{A}(2,6)s(2,6)+ N^{gsT}{}_{A}(3,6)s(3,6) + N^{gsT}{}_{A}(4,6)s(4,6) +N^{gsT}{}_{A}(5,7)s(5,7) + N^{gsT}{}_{A}(6,6)s(6,6)m^{pgT}{}_{A}(t)$$

A similar moment generating function can be obtained for the number of points in an advantage set.

Let $M^{ps}{}_{A}$, $U^{ps}{}_{A}$, $C^{ps}{}_{A}$, $S^{ps}{}_{A}$ and $K^{ps}{}_{A}$ represent the mean, standard deviation, and coefficientsof variation, skewness and kurtosis for the number of points in an advantage set. Let $M^{psT}{}_{A}$, $U^{psT}{}_{A}$, $C^{psT}{}_{A}$, $S^{psT}{}_{A}$ and $K^{psT}{}_{A}$ represent the mean, standard deviation, and coefficients of variation, skewness and kurtosis for the number of points in a tiebreaker set. Table 2 represents $M^{ps}{}_{A}$, $U^{ps}{}_{A}$, $C^{ps}{}_{A}$, $S^{ps}{}_{A}$, $K^{ps}{}_{A}$, $M^{psT}{}_{A}$, $U^{psT}{}_{A}$, $C^{psT}{}_{A}$, $S^{psT}{}_{A}$ and $K^{psT}{}_{A}$ for different values of p_A and p_B. The table covers values in the interval $0.50 \leq pA \leq pB \leq 0.75$ as this is the main area of interest for men’s tennis. It can be observed that: $M^{ps}{}_{A} > M^{psT}{}_{A}$, $U^{ps}{}_{A} > U^{psT}{}_{A}$, $C^{ps}{}_{A} > C^{psT}{}_{A}$, $S^{ps}{}_{A} > S^{psT}{}_{A}$ and $K^{ps}{}_{A} > K^{psT}{}_{A}$.

The mean number of points in a set is affected by the mean number of points in a game and the mean number of games in a set. The mean number of points in a game is greatest when p_A or $p_B = 0.50$. For a tiebreaker set, when $p_A = p_B = 0.50$, $M^{pg}{}_{A} = M^{pg}{}_{B} = 6.75$, $M^{gsT}{}_{A} = 9.66$ and $M^{psT}{}_{A} = 65.83$. When $p_A = p_B = 0.70$, $M^{pg}{}_{A} = M^{pg}{}_{B} = 5.83$, $M^{gsT}{}_{A} = 10.94$ and $M^{psT}{}_{A} = 66.22$. For this latter case, even though the mean length of games is shorter, the mean number of points in a tiebreaker set overall is greater since more games are expected to be played. Both players have a 0.90 probability of holding serve, which means that very few breaks of serve will occur and there is a 0.38 probability of reaching a tiebreaker. This is further exemplified in an advantage set, where for $p_A = p_B = 0.70$, $M^{ps}{}_{A} = 86.43$. This is also highlighted by the coefficients of variation, skewness and kurtosis being much greater for an advantage set, compared to a tiebreaker set, when p_A and p_B are both “large”.

3.3.2 APPROXIMATING THE PARAMETERS OF DISTRIBUTIONS OF THE NUMBER OF POINTS IN A SET

The moment generating function for the number of points in an advantage set $m^{ps}{}_{A}(t)$, when $p_A = 1 - p_B$, becomes:

$m^{ps}{}_A(t)=[f^{gs}{}_A(6)](m^{pg}{}_{AB})^6+[f^{gs}{}_A(7)](m^{pg}{}_{AB})^7+[f^{gs}{}_A(8)](m^{pg}{}_{AB})^8+[f^{gs}{}_A(9)](m^{pg}{}_{AB})^9+$
$[f^{gs}{}_A(10)](m^{pg}{}_{AB})^{10}+N^{gs}{}_A(5,5)(1-N^{gs}{}_A(1,1))(m^{pg}{}_{AB})^{12} / [1-N^{gs}{}_A(1,1)(m^{pg}{}_{AB})^2]$

where: $m^{pg}{}_{AB}(t) = [m^{pg}{}_A(t)+m^{pg}{}_B(t)]/2$ is the average (in this case equal) of two moment generating functions.

Table 2: The parameters of the distributions of points in a tiebreaker and advantage set for different values of p_A and p_B

p_A	p_B	$M^{psT}{}_A$	$U^{psT}{}_A$	$C^{psT}{}_A$	$S^{psT}{}_A$	$K^{psT}{}_A$	$M^{ps}{}_A$	$U^{ps}{}_A$	$C^{ps}{}_A$	$S^{ps}{}_A$	$K^{ps}{}_A$
0.50	0.50	65.83	16.54	0.25	0.55	-0.04	67.71	21.15	0.31	1.62	4.75
0.50	0.60	61.99	15.97	0.26	0.65	0.18	63.39	19.76	0.31	1.71	5.39
0.50	0.70	54.73	13.75	0.25	0.85	0.88	55.39	16.01	0.29	1.89	7.46
0.50	0.75	51.64	12.44	0.24	0.89	1.20	52.06	14.08	0.27	1.92	8.44
0.60	0.60	65.59	16.03	0.24	0.55	-0.18	69.32	24.92	0.36	2.12	7.27
0.60	0.70	63.08	14.99	0.24	0.58	-0.15	68.35	27.97	0.41	2.60	10.22
0.60	0.75	60.67	14.32	0.24	0.63	-0.05	66.01	28.12	0.43	2.83	11.98
0.70	0.70	66.22	14.96	0.23	0.25	-0.81	86.43	53.11	0.61	2.47	8.67
0.75	0.75	67.59	13.74	0.20	-0.15	-0.82	125.50	101.81	0.81	2.24	7.22

Taking the natural logarithm of the moment generating function gives an alternative generating function known as the cumulant generating function. Let $\kappa^{pg}{}_A(t)=\ln[m^{pg}{}_A(t)]$ represent the cumulant generating function for the number of points in a game. This relationship can be inverted to give $m^{pg}{}_A(t) = exp(\kappa^{pg}{}_A(t))$.

The moment generating function, $m^{ps}{}_A(t)$, can be written as:

$m^{ps}{}_A(t)=f^{gs}{}_A(6)exp(6\kappa^{pg}{}_{AB}(t))+f^{gs}{}_A(7)exp(7\kappa^{pg}{}_{AB}(t))+f^{gs}{}_A(8)exp(8\kappa^{pg}{}_{AB}(t))+f^{gs}{}_A(9)$
$exp(9\kappa^{pg}{}_{AB}(t))+f^{gs}{}_A(10)exp(10\kappa^{pg}{}_{AB}(t))+N^{gs}{}_A(5,5)exp(12\kappa^{pg}{}_{AB}(t))[1-N^{gs}{}_A(1,1)]/[1-N^{gs}{}_A(1,1)$
$exp(2\kappa^{pg}{}_{AB}(t))]$, when $p_A = 1 - p_B$

where: $\kappa^{pg}{}_{AB}(t) = [\kappa^{pg}{}_A(t)+ \kappa^{pg}{}_B(t)]/2$ is the average (in this case equal) of two cumulant generating functions.

This can be expressed as:

$$m^{ps}{}_A(t) = m^{gs}{}_A(\kappa^{pg}{}_{AB}(t)) \quad (1)$$

Similarly, the following result is established for $m^{psT}{}_A(t)$, when $p_A = 1 - p_B$:

$$m^{psT}{}_A(t)=m^{gsT}{}_A(\kappa^{pg}{}_{AB}(t))+N^{gsT}{}_A(6,6)exp(12\kappa^{pg}{}_{AB}(t))(exp(\kappa^{pgT}{}_{AB}(t))-exp(\kappa^{pg}{}_{AB}(t))) \quad (2)$$

Notice the last term does not vanish due to the difference in the scoring system for a tiebreaker game compared with a regular game. Equations (1) and (2) can be used to obtain approximate results for the parameters of distributions for the number of points in a set, when p_A is not equal to 1 - p_B.

3.4 THE NUMBER OF POINTS IN A MATCH

From this point an advantage match is considered as a match where the first four sets played are tiebreaker sets and the fifth set is an advantage set.

The moment generating functions for the number of points in an advantage and tiebreaker match, $m^{pm}(t)$ and $m^{pmT}(t)$, when $p_A = 1 - p_B$ become:

$m^{pmT}(t) = m^{sm}(\kappa^{psT}{}_{AB}\,(t))$

$m^{pm}(t) = m^{sm}(\kappa^{psT}{}_{AB}\,(t)) + N^{sm}(2{,}2)\ exp(4\kappa^{psT}{}_{AB}\,(t))\,(exp(\kappa^{ps}{}_{AB}\,(t)) - exp(\kappa^{psT}{}_{AB}\,(t)))$

where: $\kappa^{psT}{}_{AB}\,(t) = [\kappa^{psT}{}_{A}\,(t) + \kappa^{psT}{}_{B}\,(t)] / 2$ and $\kappa^{ps}{}_{AB}\,(t) = [\kappa^{ps}{}_{A}\,(t) + \kappa^{ps}{}_{B}\,(t)] / 2$

The following approximation results can be established for the number of points in a match, similar to the approximation results established for the number of points in a set:

$m^{pmT}(t) \approx m^{sm}(\kappa^{psT}{}_{AB}\,(t))$ for all values of p_A and p_B.

$m^{pm}(t) \approx m^{sm}(\kappa^{psT}{}_{AB}\,(t)) + N^{sm}(2{,}2)\ exp(4\kappa^{psT}{}_{AB}\,(t))\ (exp(\kappa^{ps}{}_{AB}\,(t)) - exp(\kappa^{psT}{}_{AB}\,(t)))$ for all values of p_A and p_B.

Approximation results for distributions of points in a match, could also be established for tennis doubles by using the above results established for singles. The probability of a team winning a point on serve is estimated by the averages of the two players in the team.

When $p_A = 1\text{-}p_B$, the distribution of number of points played each set if player A serves first in the set, is equal to the number of points played each set if player B serves first in the set. This leads to the following result:

The number of points played each set in a match are independent, if $p_A = 1 - p_B$.

Suppose $Z = X + Y$, where X and Y are independent, then it is well known that $m_Z(t) = E[e^{Zt}] = E[e^{Xt}]E[e^{Y\,t}] = m_X(t)m_Y\,(t)$. By taking logarithms it follows that $\kappa_Z(t) = \kappa_X(t) + \kappa_Y\,(t)$.

An extension of this property of cumulants is given by the following theory (Brown, 1977) and can be applied to points in a tiebreaker match when the number of points played each set in a match are independent. When the independence assumption fails to hold the theory remains approximately correct according to the approximation result established for points in a tiebreaker match.

Theorem

If $Z = X_1 + X_2 + \ldots\ldots\ldots + X_N$ where X_i are *i.i.d.* then $\kappa_Z(t) = \kappa_N(\kappa_X(t))$
Taking the derivatives of the result and setting $t = 0$ gives the following useful results in terms of cumulants:

$k^{(1)}{}_Z = k^{(1)}{}_N \, k^{(1)}{}_X$
$k^{(2)}{}_Z = k^{(2)}{}_N \, [k^{(1)}{}_X\,]^2 + k^{(1)}{}_N \, k^{(2)}{}_X$
$k^{(3)}{}_Z = 3k^{(1)}{}_X \, k^{(2)}{}_N \, k^{(2)}{}_X + [k^{(1)}{}_X\,]^3 k^{(3)}{}_N + k^{(1)}{}_N \, k^{(3)}{}_X$
$k^{(4)}{}_Z = 3k^{(2)}{}_N \, [k^{(2)}{}_X\,]^2 + 6[k^{(1)}{}_X\,]^2 k^{(2)}{}_X \, k^{(3)}{}_N + 4k^{(1)}{}_X \, k^{(2)}{}_N \, k^{(3)}{}_X + [k^{(1)}{}_X\,]^4 k^{(4)}{}_N + k^{(1)}{}_N \, k^{(4)}{}_X$

For example the mean number of points in a tiebreaker match, M^{pmT}, with the associated variance, V^{pmT}, can be calculated from the cumulant generating function as:

$$M^{pmT} = M^{psT} M^{sm}$$

$$V^{pmT} = V^{sm}(M^{psT})^2 + M^{sm} V^{psT}$$

where:
M^{psT} represents the mean number of points in a tiebreaker set
M^{sm} represents the mean number of sets in a tiebreaker match
V^{psT} represents the variance of the number of points in a tiebreaker set
V^{sm} represents the variance of the number of sets in a tiebreaker match

Let M^{pm} , U^{pm} , C^{pm} , S^{pm} and K^{pm} represent the mean, standard deviation, and coefficients of variation, skewness and kurtosis for the number of points in an advantage match. Let M^{pmT} , U^{pmT} , C^{pmT} , S^{pmT} and K^{pmT} represent the mean, standard deviation, and coefficients of variation, skewness and kurtosis for the number of points in a tiebreaker match. Tables 3 and 4 represent the exact parameters of the distributions for an advantage and tiebreaker match for different values of p_A and p_B. The results agree with Pollard (1983) for a best-of-three sets tiebreaker match. It shows that the mean, standard deviation, coefficients of variation, skewness and kurtosis of the number of points played are greater for an advantage match, compared to a tiebreaker match. Also included in the tables are the probabilities of the match lasting for at least n points, represented by $P(n)$ for an advantage match and $Q(n)$ for a tiebreaker match. These probabilities were calculated using the NP-expansion technique (Pesonen, 1987). Notice that when p_A and p_B become "large", the probability of playing at least 400 points in an advantage match is considerably greater than for a tiebreaker match. This is some justification as to why an advantage match can seemingly never end with two strong servers.

Table 5 represents the exact parameters of distributions for a tiebreaker and an advantage match using 50-40 games, along with the probability of a match going beyond 300 points. For an extreme case, when $p_A = p_B = 0.75$, the probability of an advantage match going beyond 300 points is 0.06. In comparison to Tables 3 and 4, the probability of an advantage or tiebreaker match going beyond 300 points is 0.38. This shows that replacing standard 'deuce' games with 50-40 games, substantially decreases the likelihood of long matches occurring.

It is often the case that by shortening the length of matches, decreases the probability of winning for the better player. However this is not necessarily the case as shown by replacing standard 'deuce' games with 50-40 games. Table 6 represents the probabilities of winning under four different scoring systems, for different values of p_A and p_B.

Notice when p_A=0.75 and p_B = 0.70, the probability of player A (the stronger player) winning using 50-40 games is greater than using standard 'deuce' games.

Table 3: The parameters of the distributions of points in an advantage match for different values of p_A and p_B

p_A	p_B	M^{pm}	U^{pm}	C^{pm}	S^{pm}	K^{pm}	$P(300)$	$P(350)$	$P(400)$
0.50	0.50	272.27	62.58	0.23	0.13	-0.54	0.33	0.12	0.02
0.55	0.55	272.10	62.55	0.23	0.15	-0.51	0.33	0.12	0.02
0.60	0.60	271.96	62.85	0.23	0.21	-0.37	0.33	0.12	0.02
0.65	0.65	273.48	65.13	0.24	0.40	0.14	0.32	0.12	0.03
0.70	0.70	280.72	74.16	0.26	0.92	1.96	0.34	0.15	0.06
0.75	0.75	300.52	103.49	0.34	1.89	6.23	0.38	0.22	0.12

Table 4: The parameters of the distributions of points in a tiebreaker match for different values of p_A and p_B

p_A	p_B	M^{pmT}	U^{pmT}	C^{pmT}	S^{pmT}	K^{pmT}	$Q(300)$	$Q(350)$	$Q(400)$
0.50	0.50	271.56	61.40	0.23	0.06	-0.67	0.33	0.11	0.02
0.55	0.55	271.25	61.12	0.23	0.06	-0.67	0.33	0.11	0.02
0.60	0.60	270.56	60.42	0.22	0.06	-0.69	0.32	0.10	0.01
0.65	0.65	270.52	59.77	0.22	0.05	-0.73	0.32	0.10	0.01
0.70	0.70	273.14	59.64	0.22	0.02	-0.79	0.34	0.11	0.01
0.75	0.75	278.81	59.54	0.21	-0.04	-0.88	0.38	0.13	0.02

Table 5: The parameters of the distributions of points in a tiebreaker and advantage match using 50-40 games for different values of p_A and p_B

p_A	p_B	M^{pmT}	U^{pmT}	C^{pmT}	S^{pmT}	K^{pmT}	$Q(300)$	M^{pm}	U^{pm}	C^{pm}	S^{pm}	K^{pm}	$P(300)$
0.50	0.50	198.43	44.01	0.22	0.05	-0.71	0.01	198.94	44.93	0.23	0.14	-0.54	0.02
0.55	0.55	199.71	44.39	0.22	0.05	-0.71	0.02	200.09	45.08	0.23	0.11	-0.58	0.02
0.60	0.60	201.93	44.89	0.22	0.05	-0.71	0.02	202.31	45.58	0.23	0.11	-0.58	0.02
0.65	0.65	205.18	45.52	0.22	0.05	-0.71	0.02	205.71	46.47	0.23	0.14	-0.53	0.03
0.70	0.70	209.79	46.40	0.22	0.06	-0.71	0.03	210.71	48.10	0.23	0.21	-0.38	0.04
0.75	0.75	216.64	47.81	0.22	0.06	-0.73	0.05	218.71	51.70	0.24	0.39	0.12	0.06

Table 6: The probabilities of winning a tennis match under different scoring systems standard games 50-40 games

		standard games		50-40 games	
p_A	p_B	p^m	p^{mT}	p^m	p^{mT}
0.51	0.50	0.567	0.567	0.554	0.554
0.55	0.50	0.800	0.799	0.754	0.754
0.60	0.50	0.952	0.951	0.918	0.917
0.61	0.60	0.565	0.564	0.557	0.557
0.65	0.60	0.789	0.785	0.764	0.763
0.70	0.60	0.941	0.938	0.927	0.926
0.71	0.70	0.560	0.558	0.559	0.559
0.75	0.70	0.772	0.760	0.775	0.772

4. CONCLUSION

The mathematical methods of generating functions have been used to calculate the parameters of distributions of the number of points in a tennis match. The results show that the likelihood of long matches can be substantially reduced by using the tiebreak game in the fifth set, or more effectively by using the 50-40 game throughout the match.

We used the number of points played in a match as a measure of its length. This measure is related to the time duration of the match and avoids the complications of delays between points, at change of serve, at change of end, injury time and weather delays. Further work could involve calculating the time duration of a match from the results presented in Subsection 3.4. This could then be used to calculate the probabilities of the match going beyond a given amount of time. This would provide commentators and tournament officials with very useful information on when the match is going to finish.

5. REFERENCES

Barnett, T. and Clarke, S.R. (2005) Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*. **16**(2), 113-120.

Brown, A. (1977) Cumulants of convolution-mixed distributions. *Astin Bulletin*. **9**(1,2), 59-63. Croucher, J.S. (1986) The conditional probability of winning games of tennis. *Research Quarterly for Exercise and Sport*. 57(1), 23-26.

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

Pollard, G.H. (1983) An analysis of classical and tie-breaker tennis. *Australian Journal of Statistics*. **25**(3), 496-505.

Pollard, G.H. and Noble, K. (2002) The characteristics of some new scoring systems in tennis. *In Proceedings of the 6M&CS*. G. Cohen ed., 221-226.

Pollard, G.H. and Noble, K. (2004) The benefits of a new game scoring system in tennis the 50-40 game. *In Proceedings of the 7M&CS*. R.H. Morton and S. Ganesalingam eds., 262-265.

Stuart, A. and Ord, J.K. (1987) *Kendall's Advanced Theory of Statistics*. Charles Griffin & Company Limited. London, **1**(5).