

# The relationship of excitement, importance and entropy to the efficiency of statistical hypothesis tests

Graham Pollard\*

*Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, Australia*

**Abstract.** Sports scoring systems can be seen as statistical hypothesis tests, and the tennis scoring system in particular can be viewed as a particular sequential 2-sample binomial test. Thus, the excitement of a point (or observation) within a scoring system can be seen to have a place within hypothesis testing. The 2-sample binomial tests known to be the most efficient are shown to have related characteristics with respect to the measures of excitement and importance in sport and with Shannon entropy in information theory. This paper also includes several new, interesting and quite general scoring systems theorems.

**Keywords:** Sports scoring systems, excitement of a point, importance of a point, entropy of a point, efficiency of a scoring system, efficiency of a hypothesis test

## 1. Introduction

Miles (1984), in his very elegant paper, showed that *a scoring system in sport could be considered to be identical to a hypothesis test in statistics*. In particular, he considered tennis which is an application of comparing two binomial probabilities. In tennis there are essentially two types of points, a-points (player A serves) and b-points (B serves), as described for example in Pollard (1983). Thus, tennis scoring is a ‘bi-points’ scoring system. Player A is the better player if  $p_a > p_b$ , where  $p_a$  is player A’s probability of winning a point on his service and  $p_b$  is player B’s probability of winning a point on his service. Assuming  $p_a$  and  $p_b$  are unknowns and the objective of playing a match is to efficiently identify the better player by reducing as much as possible the error probability (the better player loses or worse player

wins) for a given expected number of points played, Miles set up symmetrical statistical hypotheses, for  $0 < p_b' < p_a' < 1$ ,

Ho:  $(p_a, p_b) = (p_a', p_b')$  (or A better than B) and

H1:  $(p_a, p_b) = (p_b', p_a')$  (or B better than A),

and considered ‘fair’ bipoints scoring systems with sensible properties. He called these scoring systems ‘biformats’. Thus, a scoring system is a statistical test, and a biformat is a statistical test of the above hypothesis in which the two types of error are equal, i.e.  $\alpha = \beta$ . Thus, the tennis scoring system is essentially a statistical test of the above hypothesis.

Miles (1984) also set up unipoints for the situation where serving is essentially neither an advantage or a disadvantage, as is typically the situation in squash. In unipoints, player A is the better player if  $p > 0.5$ , where  $p$  is player A’s probability of winning a point.

Returning to bipoints, Miles (1984) used ‘point-pairs’ (consisting of an a-point and a b-point) to set up a ‘standard family of biformats’  $W_n(\text{point-pairs})$  to represent a scoring system in which the winner is the first player to win  $n$  more point-pairs than his

---

\*Corresponding author: Graham Pollard, Faculty of Education, Science, Technology and Mathematics, University of Canberra, 2/38 Girrahween Street, Braddon, ACT 2612, Canberra, Australia. Tel.: +61 2 62572311; E-mail: grahampollard@gmail.com.

opponent (or equivalently  $2n$  points). This standard family of biformats was given an efficiency of unity. He showed that the efficiency  $\rho$  of a general biformat with key characteristics  $P$  and  $\mu$  relative to this standard family is given by

$$\rho = (2(P - Q) \ln(P/Q)) / (\mu(pa - pb) \ln(paqb/pbqa)) \quad (1)$$

where  $qa = 1 - pa$ ,  $qb = 1 - pb$ ,  $P$  is the probability player A wins,  $Q = 1 - P$ , and  $\mu$  is the expected number of points played. Thus, for given  $(pa, pb)$ , the efficiency is a function of  $P$  and  $\mu$ .

A way to thinking about the efficiency of scoring systems is this ... if you are given two scoring systems with the same value of  $\mu$  for a given pair  $(pa, pb)$ , then the more efficient scoring system at  $(pa, pb)$  is the one with the larger value of  $P$  (and hence the larger value of  $P/Q$ ).

Miles (1984) setup families of scoring systems based on alternating (AL), play-the-winner (PW) and play-the-loser (PL) service exchange mechanisms, and which used the  $W_n$  stopping rule in which the winner is the first player to be  $n$  points ahead of his opponent. Play-the-winner, for example, is the service exchange mechanism in which a point won by player A is followed by an a-point, and a point won by player B is followed by a b-point. Thus, he set up the scoring systems  $W_1(W_nAL^a, W_nAL^b)$  ( $n = 1, 2, 3, \dots$ ) which consist of two parts. Play starts with an a-point (player A serves), and service alternates until one player is  $n$  points ahead, when that part of play ceases and the second part of the scoring system begins with a b-point, and points again alternate until one player is  $n$  points ahead. If this player is the same player as the one who was  $n$  points ahead on the first part of the play, then that player is the winner. Otherwise, the scoring system starts over again. Although this seems complicated, it isn't. The two parts of the scoring system are needed to create fairness. Miles (1984) showed that this scoring system had unit efficiency, and Pollard (1986, 1992) noted that it was in fact stochastically identical to  $W_n$ (point-pairs). Using the same two-part structure but with the play-the-loser service exchange mechanism in which a point won by player A is followed by a b-point and a point won by B is followed by an a-point, Miles (1984) set up the scoring systems  $W_1(W_nPL^a, W_nPL^b)$  ( $n = 2, 3, 4, \dots$ ). He showed that they have efficiencies very slightly greater than unity when  $pa + pb > 1$ , and correspondingly that the play-the-winner (PW) scoring systems  $W_1(W_nPW^a, W_nPW^b)$  ( $n = 2,$

$3, 4, \dots$ ) have efficiencies less than unity when  $pa + pb > 1$ . By identifying  $(n-2)$  "partial-PL"  $W_n$  scoring systems "between"  $W_1(W_nPL^a, W_nPL^b)$  and  $W_1(W_nAL^a, W_nAL^b)$  when  $n = 3, 4, 5, \dots$ , Pollard (1986, 1992) showed that the (full) PL family of scoring systems  $W_1(W_nPL^a, W_nPL^b)$  was in fact the *most efficient scoring system possible* (i.e. was *optimally efficient*) when  $pa + pb > 1$  (i.e. the tennis context, for example, where serving is an advantage). The corresponding PW family  $W_1(W_nPW^a, W_nPW^b)$  ( $n = 2, 3, 4, \dots$ ) is optimal when  $pa + pb < 1$  (i.e. the volleyball context, for example, where serving is a disadvantage).

Pollard (1986, 1992) noted that the above optimal biformats for the 2-sample binomial test when  $\alpha = \beta$  (i.e. the symmetric case) are such that the biformats (or tests) pass through non-symmetrical situations and yet the 1-step PL (or PW) rule remains optimal. These non-symmetrical situations, as a starting point, correspond to the  $\alpha \neq \beta$  case in hypothesis testing and the handicap situation in sport. Thus, using the principle of optimality (dynamic programming), he concluded that the PL (PW) rule used with  $W(n1, n2)$  systems is optimally efficient for the 2-sample binomial statistical test when  $\alpha \neq \beta$  provided  $pa + pb > 1$  ( $pa + pb < 1$ ). Note that here  $W(n1, n2)$  represents the stopping rule for the handicap sporting situation in which player A needs to win  $n1$  more points than player B in order to win, and player B needs to win  $n2$  more points than player A in order win.

It has remained a challenge to describe why these PL (PW) rules lead to optimal tests when used with  $W_n$  stopping rules. Further, it would appear that a greater insight into why such tests are very efficient or optimal might give researchers an increased capacity in designing very good statistical tests for other situations. Pollard (1986, pp.30 and 1992, pp.282) gave an intuitive explanation for the super-efficiency of these PL and PW scoring systems. It was that 'in each case, under the most efficient rule, a greater number of points with the more variable (Bernoulli) distribution is expected to be selected; and this is an approach used in efficient estimation theory'. Pollard (1986, 1990) gave an alternative explanation for the case in which the value of  $n$  in the above PL and PW scoring systems is very large, and this explanation is outlined in the following paragraph.

Pollard (1990, pp.191) described 'a method for decomposing many sequential probability ratio [statistical] tests [including the PL and PW systems described above] into smaller independent

components called “modules”. A function of some characteristics of modules can be used to determine the asymptotically most efficient of a set of statistical tests...’, and he ‘used the “module” method to give an explanation for the super-efficiency of the play-the-winner (when  $p_a + p_b < 1$ ) and play-the-loser rules (when  $p_a + p_b > 1$ ) in two sample binomial sampling’. In the context of  $W_m$  systems and denoting player A as the better player, he defined the random variable  $S$  as the increase (or shift) in player A’s score during one module and the random variable  $D$  as the duration (i.e. number of points played) of the module. He compared  $W1(W_mPL^a, W_mPL^b)$  and  $W1(W_mPW^a, W_mPW^b)$  when  $p = (p_a + p_b)/2$ ,  $q = 1 - p$ ,  $m = nq$  for the PL case,  $m = np$  for the PW case, and  $n$  is large, noting that these two systems have the same asymptotic durations. The value of the variance of  $S$ ,  $Var(S)$  for the PL module was then ‘scaled’ so that it can be *directly* compared with the value of  $Var(S)$  for the PW module. The purpose of this scaling was to allow for (i) the different values of  $E(D)$  for the two modules, and (ii) the different values of  $m$  in  $W_m$  for the two systems. He concluded (pp.197) that ‘for the same asymptotic expected durations [of the two  $W_m$  systems], the smaller scaled (or “relative”) variance for the PL module (when  $p_a \neq p_b$  and  $p_a + p_b > 1$ ) results in the better player having a higher probability of winning and hence the PL system has greater efficiency’. Using this ‘module’ approach to the asymptotic case he also concluded that if  $p_a + p_b < 1$  and  $p_a \neq p_b$ , the PW structure is the more efficient.

The major aim of this paper is to give further insights into why the PL or the PW structure, when combined with the  $W_n$  stopping rule, produces optimally efficient scoring systems. Unlike the results described in the previous paragraph, in this paper we do not restrict consideration to the situations in which the  $n$  in  $W_n$  is very large. We gain insight into this by considering the ‘importance’ of a point within a scoring system, the ‘excitement’ of a point within a scoring system, and the ‘information’ or ‘entropy’ of a Bernoulli trial. These concepts are described in the following paragraphs.

In an elegant paper Morris (1977) defined the ‘importance’ of a point within a game as the probability player A wins the game given he wins that point minus the probability he wins the game given he loses the point. Pollard (1986, 1992) noted that the very efficient  $W_n$  scoring systems (using AL, PL and PW) had equally important points when the players were equal ( $p_a = p_b$ ).

In information theory, Shannon (1948) defined the entropy of a Bernoulli observation with probability  $p$  to be equal to  $p \ln(1/p) + (1 - p) \ln(1/(1 - p))$ .

In the paper by Pollard (2016) a mathematical measure for the excitement of a point in various sports was devised. In this paper we give a second excitement measure, similar in concept but one based on squared values rather than absolute values.

Pollard (1986, 1992) considered two equal players and found an expression for the increased probability player A has of winning a game given that he increases his probability of winning the point in state  $i$  by  $\delta_i$  ( $i = 1, 2, 3, \dots$ ). He showed that player A’s increased probability of winning the game is given by  $\Delta P = \sum N_i * I_i * \delta_i$ , where  $I_i$  is the importance of the point at state  $i$  when the players are equal and  $N_i$  is the expected number of times state  $i$  is entered in one realization of the game with player A’s increased point probabilities. An extension of this result is given in Section 2. This paper also includes several new and quite general scoring systems theorems, including a second relationship between player A’s probability of winning under a scoring system and the importances of the points within that system. Indeed, these theorems are of general interest in their own right.

## 2. Method

Pollard (2016) noted that there are several factors that can contribute to the excitement of a situation in sport. Indeed, he focused on a mathematical approach based on outcome probabilities for the point. He noted that the excitement of a particular point within a scoring system can be measured by the expected value of the absolute size of the change in a player’s probability of an overall win as a result of that point being played. It follows that the excitement of a point is related to the importance of the point (Morris (1976)). Namely, if  $p$  is player A’s probability of winning a point with importance  $I$ , the excitement of that point is equal to  $2pqI$ , where  $q = 1 - p$ .

We begin by considering two simple and very similar scoring systems that are particularly relevant to this paper. They are  $W2AL^a$  and  $W2AL^b$  when  $p_a = 0.9$  and  $p_b = 0.5$ . It can be shown that when the score is  $-1$  (player A trailing by 1 point), player A’s probability of winning is 0.81 under  $W2AL^b$  and that it is 0.45 under  $W2AL^a$ , whilst his probability of winning at the score of 0 is 0.9 under both systems, and it is clearly zero at the score of  $-2$  for both systems.

Thus, the point played at the score of  $-1$  has an importance of  $0.9$  under both scoring systems. However, the point played at the score of  $-1$  is more exciting under scoring system  $W2AL^a$  than it is under the scoring system  $W2AL^b$ . The expectation of the absolute value of the change in player A's probability of winning the game at the score of  $-1$  is equal to  $0.5*|0.9 - 0.45| + 0.5*|0.45 - 0| = 0.45$  for  $W2AL^a$ , whereas it is  $0.9*|0.9 - 0.81| + 0.1*|0.81 - 0.0| = 0.162$  for  $W2AL^b$ . The excitement at the score of  $-1$  is greater under  $W2AL^a$  because player A's (and hence also player B's) probability of winning is about to experience a greater expected absolute change under that system. This example highlights the difference between the importance of a point and the excitement of a point. It is easy to show that the excitement of a point is in fact equal to the importance of that point multiplied by  $2pq$  where  $p$  is the probability that player A wins the point, and  $q = 1 - p$ .

Continuing briefly with the above example, for  $W2AL^a$  the excitement at the score of  $+1$  can be shown to be  $0.05$  and the excitement at the score of  $0$  to be  $0.09$ . Thus, starting from a score of  $0$ , the expected total excitement for the first two points is equal to  $1*0.09 + 0.9*0.05 + 0.1*0.45 = 0.18$ . For the system  $W2AL^b$  the excitement at the score of  $+1$  is equal to  $0.018$  and the excitement at the score of  $0$  is equal to  $0.09$ . Thus, starting from a score of  $0$  in  $W2AL^b$ , the expected total excitement for the first two points is equal to  $1*0.09 + 0.5*0.018 + 0.5*0.162 = 0.18$ , interestingly the same as for  $W2AL^a$ .

In the above paragraph the excitement of a point has been defined as the expectation of the absolute value of the change in player A's probability of winning the game as a result of the playing of that point. It can be seen that this is a measure of 'probability variation' resulting from the playing of that point. As in other areas of statistics it is possible to devise other measures of variation, and the squared measure is a standard one. Thus, we might define a second measure of excitement based on the square of changes in probability. The second measure of excitement is the expected value of the square of the change in player A's probability of winning as a result of the playing of that point. This measure is called Ex2, and the first measure Ex1. It is straightforward to show that Ex2 is equal to  $pqI^2$  where  $I$  is the importance of the point,  $p$  is the probability player A wins the point, and  $q = 1 - p$ . We note here that, just as the measure Ex1 can be used in the situation where there are more than just the two outcomes, win or loss, this second measure

Ex2 can also be used where there are three or more outcomes (e.g. win, draw or loss).

Four new and quite general scoring system theorems that are needed for this study are now presented.

**Theorem 1.** Consider a scoring system in which player A's probability of winning is  $P$  and the importance of the point in state  $i$  is  $I_i$ , when player A has a probability  $p_i$  of winning the point in state  $i$  and a probability  $q_i (= 1 - p_i)$  of losing the point ( $i = 1, 2, \dots$ ). Suppose he modifies his probabilities of winning the points in states  $i$  by an amount  $\delta_i$  ( $i = 1, 2, \dots$ ). Then his overall probability of winning is modified by an amount  $\Delta P = \sum n_i * I_i * \delta_i$ , where  $n_i$  is the expected number of times state  $i$  is entered in one realization of the scoring system with player A's modified point probabilities  $p_i + \delta_i$ . Note that  $I_i$  is the importance of the point in state  $i$  before the probabilities are modified.

**Proof.** We note firstly (before any modifications in player A's point probabilities) that by playing the point in state  $i$ , player A's overall probability of winning increases by  $q_i * I_i$  with probability  $p_i$  or decreases by  $p_i * I_i$  with probability  $q_i$ . This increase or decrease in Player A's probability of winning can be considered a positive or negative step (resulting from the playing of this point) along the  $P$  axis which goes from  $0$  (A loses) to  $1$  (A wins). As the expected number of times state  $i$  is entered is  $n_i$  after he modifies his point probabilities by  $\delta_i$ , the expected value of the sum of these step values is equal to  $\sum n_i * ((p_i + \delta_i) * q_i * I_i + (q_i - \delta_i) * (-p_i * I_i))$ , and this equals  $\sum n_i * \delta_i * I_i$ . The sum of these steps is also given by  $P'(1 - P) - Q'P$  where  $P'$  is player A's overall probability of winning with the modified point probabilities, and  $Q' = 1 - P'$ . Further,  $P'(1 - P) - Q'P = \Delta P$ , where  $\Delta P = P' - P$ , and this completes the proof.

**Corollary 1.** Consider two equal players (with  $p = 0.5$  on every point in unipoints or  $pa = pb$  in bipoints) playing a fair game (namely, one in which the probability player A wins the game is  $0.5$ , and the probability player B wins it is also  $0.5$ ). Suppose player A increases his probability of winning each point by  $\delta$ . Then, if  $P$  is now player A's probability of winning the game given his increased probability of winning each point,  $P/Q = (1 + 2\Delta P)/(1 - 2\Delta P)$ , where  $\Delta P = \sum n_i * I_i * \delta$ ,  $I_i$  is the importance of the point at state  $i$  when the players are equal and  $n_i$  is the expected number of times state  $i$  is entered in one realization of the game with player A's increased point probabilities.

**Proof.** This result follows from Theorem 1.

**Theorem 2.** For any scoring system which results in a win to player A with probability P or a loss to player A with probability Q (where  $P + Q = 1$ ), suppose the sum S is defined by  $S = \sum n_i p_i q_i I_i^2$  where the summation is over all the transient states i,  $n_i$  is the expected number of times state i is entered,  $p_i$  is the probability player A wins the point in that state,  $I_i$  is the importance of that state, and  $q_i = 1 - p_i$ . Thus S is the weighted sum of the values of Ex2 for each state. Then S is equal to PQ.

**Proof.** We use backward induction, and consider without loss of generality, the movement from any transient state 1 to transient state 2 with probability p1 or transient state 3 with probability q1, where  $q1 = 1 - p1$ . We assume the result is true for states 2 and 3. Thus, this sum from state 1 onwards is equal to

$$\begin{aligned} & 1 * p1 * q1 * I_1^2 + p1 * \sum n_j p_j q_j I_j^2 + q1 * \sum n_k p_k q_k I_k^2 \\ & = p1 * q1 * (P_2 - P_3)^2 + p1 * P_2 * Q_2 + q1 * P_3 * Q_3, \\ & \text{using an obvious notation} \\ & = p1 * q1 * (P_2 - P_3) * (Q_3 - Q_2) + p1 * P_2 * Q_2 + \\ & q1 * P_3 * Q_3 \\ & = (p1 * P_2 + q1 * P_3) * (p1 * Q_2 + q1 * Q_3) \\ & = P_1 * Q_1, \text{ using the same notation, and this completes the proof.} \end{aligned}$$

**Corollary 2.** For the scoring system in Theorem 2,  $P = (1 + \text{SQRT}(1 - 4 * S)) / 2$ ,  $Q = (1 - \text{SQRT}(1 - 4 * S)) / 2$ , and  $P / Q = (1 + \text{SQRT}(1 - 4 * S)) / (1 - \text{SQRT}(1 - 4 * S))$ .

**Proof.** This result follows from Theorem 2.

The above two theorems are applicable in a very general sense, and two examples are given in the Appendix to demonstrate this.

**Theorem 3.** Given two equal players (with  $p = 0.5$  on every point in unipoints or  $p_a = p_b$  in bipoints) playing a fair game (one in which the probability player A wins the game is 0.5 and the probability player B wins it is 0.5), suppose player A increases his probability of winning each point by  $\delta$ . Then,  $P * Q + (\delta * \sum n_i * I_i)^2 = 0.25$ , where  $I_i$  is the importance of the point at state i when the players are equal and  $n_i$  is the expected number of times state i is entered in one realization of the game with player A's increased point probabilities.

**Proof.** This result follows from Corollaries 1 and 2.

**Theorem 4.** Suppose the two systems  $W1(WkAL^a, WkAL^b)$  and  $W1(WmPW^a, WmPW^b)$  have the same

expected durations, and the same importances I for every point when  $p_a = p_b = p$ . Now suppose player A increases his probability of winning every point by  $\delta$ . Then  $P_{AL} * Q_{AL} = 0.25 - I^2 * \delta^2 * \mu_{AL}^2$  and  $P_{PW} * Q_{PW} = 0.25 - I^2 * \delta^2 * \mu_{PW}^2$  for the two systems, where  $\mu_{AL}$  and  $\mu_{PW}$  are the expected durations of the two systems with player A's increased probability of winning each point.

**Proof.** This result follows from Theorem 3.

**Corollary 3.** The corresponding equation to that in Theorem 4 also holds when the two systems are  $W1(WkAL^a, WkAL^b)$  and  $W1(WnPL^a, WnPL^b)$ , and when the two systems are  $W1(WnPL^a, WnPL^b)$  and  $W1(WmPW^a, WmPW^b)$ .

Returning to the major focus of this study, we note that Pollard (1986, 1992) showed that the scoring systems  $W1(WnSS^a, WnSS^b)$ , where SS is AL, PL or PW, all have the constant probability ratio property (cpr). That is, the ratio of the probability that player A wins in  $n + 2m$  points divided by the probability that he loses in  $n + 2m$  points ( $m = 0, 1, 2, \dots$ ) is a constant, and is equal to the value of P/Q for that scoring system. In particular, for the systems  $Wn(\text{point-pairs})$  and  $W1(WnAL^a, WnAL^b)$ , the cpr is  $P/Q = (p_a^n * q_b^n) / (p_b^n * q_a^n)$  and  $(P - Q) / \mu = (p_a - p_b) / 2n$ . The values of these expressions for the systems  $W1(WnPL^a, WnPL^b)$  and  $W1(WnPW^a, WnPW^b)$  are given in Table 1.

The expressions in Table 1 allow us to compare these efficient W-systems for their efficiency and other aspects. The efficiency of scoring system is a function of P and  $\mu$  (see equation (1)). It can also be seen as a function of  $(P - Q) / \mu$  and P/Q. It is clear that the average importance and the average excitement of points within a scoring system are typically smaller when the expected duration of a scoring system is larger. It is typically not possible to compare two scoring systems with exactly the same expected duration. However, with the systems in Table 1 we can compare two systems with the same value of  $(P - Q) / \mu$ , that is, with very similar values of  $\mu$ . We next show how this can be done.

The first two very efficient systems we compare are an AL one and a PW one. We suppose  $p_a = l / (l + m) + \delta$  and  $p_b = l / (l + m) - \delta$ , where l and m are integers, and  $\delta$  is such that  $p_a$  and  $p_b$  are legitimate probabilities. For the scoring system  $W1(Wn1PW^a, Wn1PW^b)$ , we have  $(P - Q) / \mu = \delta / (1 + (n - 1) * (2m / (l + m)))$ , and for the scoring system  $W1(Wn2AL^a, Wn2AL^b)$ , we have  $(P - Q) / \mu = \delta / n$ . Thus, the values for  $(P - Q) / \mu$  for these two scoring systems are equal

Table 1  
Expressions for the constant probability ratio and (P-Q)/μ for 3 scoring systems

System	cpr, P/Q	(P-Q)/μ
W1(WnAL <sup>a</sup> , WnAL <sup>b</sup> )	$(pa^n * qb^n) / (pb^n * qa^n)$	$(pa - pb) / 2n$
W1(WnPL <sup>a</sup> , WnPL <sup>b</sup> )	$(pa * qb^{2n-1}) / (pb * qa^{2n-1})$	$(pa - pb) / (2(1+(n-1)(pa+pb)))$
W1(WnPW <sup>a</sup> , WnPW <sup>b</sup> )	$(pa^{2n-1} * qb) / (pb^{2n-1} * qa)$	$(pa - pb) / (2(1+(n-1)(qa+qb)))$

if  $n_2 = 1 + (n_1 - 1) * 2 m / (1 + m)$ . Hence, for example, if  $l = 3$  and  $m = 1$ , these two systems have the same values for (P-Q)/μ when (n1, n2) are (3, 2), or (5, 3), or (7, 4), ... It follows from the form of equation (1) above that, for these values of l, m, n1 and n2, the ratio of the efficiencies of these two systems is simply the ratio of the natural logarithms their P/Q (or cpr) values.

The first of the specific examples in the previous paragraph ( $n_1 = 3$  and  $n_2 = 2$ ), namely W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) and W1(W3PW<sup>a</sup>, W3PW<sup>b</sup>) when  $l = 3$  and  $m = 1$ , is now considered in detail. It is known from earlier work described in the introduction that the system W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) is the more efficient of these two systems when  $pa + pb > 1$ . Standard numerical markov chain methods were used to find the expected number of points played and the probability player A wins using the scoring system W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>). This scoring system starts with an a-point played at state (0, a), and the first section of it is completed when the score reaches +2 when state W, 0b is reached (and the W2AL<sup>b</sup> section commences with the b-point) or when the score reaches -2 when the state L, 0b is reached (and the W2AL<sup>b</sup> section commences). It can be seen that there are nine transient states (or scores) in this formulation, and of course two absorbing states ('A wins' and 'A loses'). The numerical markov chain approach that was used is described for example in the text by Kemeny and Snell (1960, pp.43-49). This numerical approach was taken as it appeared that, in general for the various systems under consideration, explicit expressions for the importance of each point and the expected number of times each point is played, and thus the average weighted importance and the average weighted excitement would be either very complicated or not achievable. Note that this numerical approach required matrix inversion, and the matrix became quite large as the n in Table 1 became large.

When  $pa = pb = 0.75$  (i.e.  $l = 3$ ,  $m = 1$  and  $\delta = 0$ ) the two systems W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) and W1(W3PW<sup>a</sup>, W3PW<sup>b</sup>) can be shown to have the same expected duration of 64/3 points, and every

Table 2  
Characteristics of W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) when  $pa = pb = 0.75$

State	Probability	Importance, I <sub>i</sub>	n <sub>i</sub>	n <sub>i</sub> * I <sub>i</sub>
W, 1a	15/16	1/4	2/3	1/6
W, 0b	12/16	1/4	8/3	2/3
W, -1a	11/16	1/4	2	1/2
1, b	9/16	1/4	4	1
0, a	8/16	1/4	16/3	4/3
-1, b	5/16	1/4	4/3	1/3
L, 1a	7/16	1/4	2/3	1/6
L, 0b	4/16	1/4	8/3	2/3
L, -1a	3/16	1/4	2	1/2
Total a, Na			32/3	8/3
Total b, Nb			32/3	8/3
Total or Average			64/3	0.25

point within each of the two systems has an importance of 0.25. Table 2 gives, for the W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) system, the probability that player A wins from each of the nine transient states, the importance of each state, and n<sub>i</sub>, the number of times each state is entered in one realization of W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>).

Table 3 gives the same information as Table 2, but for the parameter values  $pa = 0.76$  and  $pb = 0.74$  (i.e.  $\delta = 0.01$ ). It also gives an additional column. The mean duration of W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) when  $pa = 0.76$  and  $pb = 0.74$  is equal to 21.2615 and the average weighted importance of the points is 0.2483. Note that values are given to 4 decimal places, unless it is considered necessary or useful to do otherwise. It follows from Table 3 that the weighted sum of the excitements Ex1 is 1.9774, and the weighted sum of the excitements Ex2 is 0.2472. Also, the average importance per point played is equal to 0.0117, the average Ex1 per point played is 0.0930, and average Ex2 per point played is 0.0116.

The analyses summarized in Tables 2 and 3 for W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) were repeated for W1(W3PW<sup>a</sup>, W3PW<sup>b</sup>), and a comparison of the two systems is given in Table 4. The efficiency of W1(W3PW<sup>a</sup>, W3PW<sup>b</sup>) when  $pa = 0.76$  and  $pb = 0.74$  is equal to 0.999822 (from equation (1)), whilst W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) has unit efficiency. It can be seen that the more efficient of these two scoring systems, W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>), has the following seven characteristics

Table 3  
Characteristics of W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) when pa=0.76 and pb=0.74

State	Probability	Importance, I <sub>i</sub>	n <sub>i</sub>	n <sub>i</sub> * I <sub>i</sub>	n <sub>i</sub> * I <sub>i</sub> <sup>2</sup>
W, 1a	0.9492	0.2115	0.7278	0.1539	0.0326
W, 0b	0.7885	0.2172	2.7994	0.6081	0.1321
W, -1a	0.7320	0.2353	2.0715	0.4875	0.1147
l, b	0.6143	0.2353	4.0397	0.9507	0.2237
0, a	0.5532	0.2549	5.3154	1.3551	0.3455
-1, b	0.3594	0.2618	1.2757	0.3340	0.0875
L, 1a	0.4903	0.2618	0.6542	0.0713	0.0448
L, 0b	0.2913	0.2689	2.5160	0.6766	0.1819
L, -1a	0.2214	0.2913	1.8619	0.5424	0.1580
Total a, Na			10.6308	2.7102	0.6956
Total b, Nb			10.6308	2.5694	0.6252
Total or Average			21.2615	0.2483	

Table 4  
A comparison of W1(W2AL<sup>a</sup>, W2AL<sup>b</sup>) and W1(W3PW<sup>a</sup>, W3PW<sup>b</sup>) when pa=0.76 and pb=0.74

	W1(W2AL <sup>a</sup> , W2AL <sup>b</sup> )	W1(W3PW <sup>a</sup> , W3PW <sup>b</sup> )
Na	10.630774	10.841475
Nb	10.630774	10.416320
Expected duration, μ	21.261549	21.257795
Probability A wins, P	0.5531539	0.5531445
Efficiency	1.0	0.999822
cpr	1.237907	1.237860
(P-Q)/μ	0.005	0.005
P*Q	0.2471747	0.2471757
S = Σ n <sub>i</sub> p <sub>i</sub> q <sub>i</sub> I <sub>i</sub> <sup>2</sup>	0.2471747	0.2471757
0.25 - I <sup>2</sup> *δ <sup>2</sup> *μ <sup>2</sup> (with I=0.25)	0.2471747	0.2471757
Average Importance, a-points	0.2549437	0.2449901
Average Importance, b-points	0.2416930	0.2516034
Average Importance, all points	0.2483183	0.2482306
Average Importance per point played	0.0116792	0.0116772
Average total excitement, Ex1, a-points	0.9886987	0.9689286
Average total excitement, Ex1, b-points	0.9886987	1.0084767
Average total excitement, Ex1, all points	1.9773973	1.9774053
Average total excitement Ex1 per point played	0.093003	0.093020
Average total excitement, Ex2, a-points	0.1268820	0.1192684
Average total excitement, Ex2, b-points	0.1202927	0.1279072
Average total excitement, Ex2, all points	0.2471747	0.2471756
Average total excitement Ex2 per point played	0.0116254	0.0116275
Average total Shannon entropy	11.9504	11.9437
Average total Shannon entropy per point played	0.56207	0.56185

1. a greater cpr,
2. points with greater average importance,
3. a greater value for the average importance per point played,
4. points with smaller average total excitements, Ex1 and Ex2,
5. smaller values for the average total excitement per point played,
6. greater average total Shannon entropy, and
7. greater average total Shannon entropy per point played.

As a second example we compare W1(W3PL<sup>a</sup>, W3PL<sup>b</sup>) with W1(W4AL<sup>a</sup>, W4AL<sup>b</sup>) when pa=0.76

and pb=0.74. The value of (P-Q)/μ for each of these two scoring systems is equal to 0.0025. Equation (1) can be used to show that the system W1(W3PL<sup>a</sup>, W3PL<sup>b</sup>) has an efficiency of 1.000089 whilst W1(W4AL<sup>a</sup>, W4AL<sup>b</sup>) has efficiency of 1 when pa=0.76 and pb=0.74. (It is noted that when pa=pb=0.75, both these scoring systems have an expected duration of 256/3 points and all points have an importance of 0.125). It was shown that the more efficient scoring system W1(W3PL<sup>a</sup>, W3PL<sup>b</sup>) has the same seven characteristics noted above.

As a third example we compare W1(W3PL<sup>a</sup>, W3PL<sup>b</sup>) with W1(W5PW<sup>a</sup>, W5PW<sup>b</sup>) when pa=2/3 + 0.01 and pb=2/3 - 0.01. The value of (P-Q)/μ

Table 5  
 $W1(W6PL^a, W6PL^b)$  and  $W1(W8AL^a, W8AL^b)$  when  $pa=0.71$  and  $pb=0.69$

	$W1(W6PL^a, W6PL^b)$	$W1(W8AL^a, W8AL^b)$
Na	143.642952	145.450714
Nb	147.279482	145.450714
Expected duration, $\mu$	290.922434	290.901428
Probability A wins, P	0.6818265	0.6818134
Efficiency	1.000079406	1.0
cpr	2.142939516	2.142809837
$(P-Q)/\mu$	0.00125	0.00125
$P*Q$	0.2169391	0.2169439
Average total excitement $Ex2 = S = \sum n_i p_i q_i I_i^2$	0.2169391	0.2169439
Average total excitement Ex2 per point played	0.000745694	0.000745764
Average total Shannon entropy	177.67567	177.63203
Average total Shannon entropy per point played	0.610732	0.610626

for each of these two scoring systems is equal to 3/1100. Equation (1) can be used to show that the system  $W1(W3PL^a, W3PL^b)$  has an efficiency of 1.0000546 and  $W1(W5PW^a, W5PW^b)$  has efficiency of 0.999891 when  $pa = 2/3 + 0.01$  and  $pb = 2/3 - 0.01$ . (It is noted that when  $pa = pb = 2/3$ , both these scoring systems have an expected duration of 30.25 and all points have an importance of 3/22.) When  $pa = 2/3 + 0.01$  and  $pb = 2/3 - 0.01$ , the more efficient scoring system  $W1(W3PL^a, W3PL^b)$  has the same seven characteristics noted in the two examples above.

A modified approach for deriving *some* of the results in the above table (but not those for Ex1) is now described. This approach is useful when the values of  $n$  in these systems are not as small as in the above examples, and the matrices needed to be inverted (numerically) would be quite large. To demonstrate this, suppose we are interested in calculating the characteristics for  $W1(WnAL^a, WnAL^b)$  and  $W1(WmPL^a, WmPL^b)$  when  $(pa, pb)$  equals, say, (0.71, 0.69). These two systems have the same value for  $(P-Q)/\mu$  when  $n=8$  and  $m=6$ . Making use of the expressions in Table 1, equation (1), and Theorem 2, the results for these two systems are shown in Table 5. It can be seen that the more efficient scoring system  $W1(W6PL^a, W6PL^b)$  has the following 5 characteristics

1. a greater cpr,
2. points with smaller total average excitement Exc2,
3. a smaller value for the total average excitement Exc2 per point played,
4. greater average total Shannon entropy, and
5. greater average total Shannon entropy per point played.

As a second example of this modified approach, suppose we are interested in calculating the charac-

teristics for  $W1(WnAL^a, WnAL^b)$  and  $W1(WmPW^a, WmPW^b)$  when  $(pa, pb)$  equals, say (0.71, 0.69). Here it can be shown that  $W1(W7AL^a, W7AL^b)$  and  $W1(W11PW^a, W11PW^b)$  have the same value of  $(P-Q)/\mu$ , and the more efficient system  $W1(W7AL^a, W7AL^b)$  can be shown to have the same five characteristics noted above.

As a third example we compare  $W1(WnAL^a, WnAL^b)$  and  $W1(WmPL^a, WmPL^b)$  when  $(pa, pb)$  equals, say (0.72, 0.70). Here  $W1(W72AL^a, W72AL^b)$  and  $W1(W51PL^a, W51PL^b)$  have the same value of  $(P-Q)/\mu$ , and the more efficient system  $W1(W51PL^a, W51PL^b)$  can be shown to have the same five characteristics noted above.

In general, it can be seen that, given any two values of  $pa$  and  $pb$ , it is possible to compare any two of the three general W systems under consideration (AL, PW, PL) by using an appropriate  $n$  and  $m$  value, and the more efficient system can be shown using the 'matrix approach' to have the seven characteristics noted above. Correspondingly, the 'non-matrix approach' can be used to show that the more efficient system has the five characteristics noted secondly.

### 3. Conclusions

It has been noted that a set of tennis can be considered to be a rather specific design for a sequential 2-sample binomial statistical test, albeit not an efficient one. The excitement of a particular point within a tennis set can be measured by the absolute change in a player's probability of winning the set as a result of that point being played. A squared measure of excitement has also been defined. These excitement measures for a point are thus measures of variability in a player's probability of winning as a result of

the point being played. The most efficient sequential 2-sample binomial test in a given situation has been shown to have the smallest average excitement per point played (or per bernouilli trial). This most efficient binomial test has also been shown to have the largest average ‘importance’ per point played and the largest ‘Shannon entropy’ per point played.

These characteristics of the most efficient 2-sample binomial tests give further insights into the design of efficient statistical tests, and may be useful conceptually in the design of statistical tests for other situations.

## References

- Kemeny, J.G., Snell, J.L., 1960. *Finite Markov Chains*. Princeton, New Jersey: D. Van Nostrand.
- Miles, R., 1984. Symmetric sequential analysis: The efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society B*. 46(1), 93–108.
- Morris, C., 1977. The most important points in tennis. In *Optimal Strategies in Sport*, edited by S. P. Ladany and R. E. Machol, 131–140. Amsterdam: North Holland. (Vol. 5 in *Studies in Management Science and Systems*.)
- Pollard, G.H., 1983. An analysis of classical and tie-breaker tennis. *Australian Journal of Statistics*. 25(4), 496–505.
- Pollard, G.H., 1986. *A stochastic analysis of scoring systems*. PhD thesis, Australian National University.
- Pollard, G.H., 1990. A method for determining the asymptotic efficiency of some sequential probability ratio tests. *Australian Journal of Statistics*. 32(2), 191–204.
- Pollard, G.H., 1992. The optimal test for selecting the greater of two binomial probabilities. *Australian Journal of Statistics*. 34, 273–284.
- Pollard, G.H., 2017. Measuring excitement in sport. *Journal of Sports Analytics*, 3(1), 37–43.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 379–423 and 623–656.

## Appendix

**Example A1.** This is an example of Theorem 2. Consider a game in which the winner is the first player to reach 3 ‘units’. Suppose player A gets 1 ‘unit’ for each point he wins, and player B gets 1 ‘unit’ for each point he wins except the first point played for which he gets two ‘units’ if he wins it. Clearly this is an unfair game. Using the notation  $(n_1, n_2, n_3)$  to represent the state where player A has won  $n_1$  ‘units’, player B has won  $n_2$  ‘units’ and  $n_3$  points have been played, the possible states in this game are  $(0,0,0), (1,0,1), (0,2,1), (2,0,2), (1,1,2), (1,2,2), (0,3,2)=L, (3,0,3)=W, (2,1,3), (2,2,3), (1,2,3), (1,3,3)=L, (3,1,4)=W, (3,2,4)=W, (2,2,4), (2,3,4)=L, (1,3,4)=L, (3,2,5)=W$  and  $(2,3,5)=L$ . The absorbing states where player A loses (L) or wins (W) are identified. Taking the 10 transient (or non-absorbing) states in the above order, and supposing the  $p$ -values (for the probability that player A wins the point in that state) are respectively 0.8, 0.3, 0.7, 0.4, 0.5, 0.6, 0.4, 0.5, 0.6 and 0.6, it can be shown

(by calculating backwards from the last possible point) that the importances of these points are equal to 0.4388, 0.296, 0.3, 0.24, 0.4, 0.5, 0.4, 1.0, 0.6 and 1.0 respectively. For these values of  $p$ , the values of  $n_i$  for these transient states are equal to 1, 0.8, 0.2, 0.24, 0.56, 0.14, 0.424, 0.084, 0.28 and 0.4224 respectively. It follows that  $\sum n_i p_i q_i I_i^2 = 0.2462741184$ . First principles can be used to show that the probability A wins the game,  $P$  is equal to 0.56104, and Theorem 2 can then be verified.

**Example A2.** We now suppose player A in Example A1 modifies his  $p$ -values in the 10 transient states in the above example by  $\delta_i$  values of  $-0.2, 0.1, -0.1, 0.1, 0.1, 0.1, 0.1, -0.1, 0.1,$  and  $-0.1$  respectively. The  $n_i$  values are now equal to 1, 0.6, 0.4, 0.24, 0.36, 0.24, 0.336, 0.168, 0.144 and 0.2688 and, using Theorem 1,  $\Delta P = \sum n_i \delta_i I_i = -0.07144$  and so player A’s probability of winning is now  $P' = 0.56104 - 0.07144 = 0.4896$ , which can be verified by first principles.