

A Stochastic Analysis of Some Two-Person Sports

By Davy Paindaveine and Yvik Swan

We consider two-person sports where each rally is initiated by a *server*, the other player (the *receiver*) becoming the server when he/she wins a rally. Historically, these sports used a scoring based on the *side-out scoring system*, in which points are only scored by the server. Recently, however, some federations have switched to the *rally-point scoring system* in which a point is scored on every rally. As various authors before us, we study how much this change affects the game. Our approach is based on a *rally-level analysis* of the process through which, besides the well-known probability distribution of the scores, we also obtain the distribution of the number of rallies. This yields a comprehensive knowledge of the process at hand, and allows for an in-depth comparison of both scoring systems. In particular, our results help to explain why the transition from one scoring system to the other has more important implications than those predicted from game-winning probabilities alone. Some of our findings are quite surprising, and unattainable through Monte Carlo experiments. Our results are of high practical relevance to international federations and local tournament organizers alike, and also open the way to efficient estimation of the rally-winning probabilities.

1. Introduction

We consider a class of two-person sports for which each rally is initiated by a *server*—the other player is then called the *receiver*—and for which the rules and scoring system satisfy one of the following two definitions.

Address for correspondence: Davy Paindaveine, Université Libre de Bruxelles, E.C.A.R.E.S. and Département de Mathématique, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Bruxelles, Belgium; e-mail: dpaindav@ulb.ac.be

Side-out scoring system: (i) The server in the first rally is determined by flipping a coin. (ii) If a rally is won by the server, the latter scores a point and serves in the next rally. Otherwise, the receiver becomes the server in the next rally, but no point is scored. (iii) The winner of the *game* is the first player to score n points.

Rally-point scoring system: (i) The server in the first rally is determined by flipping a coin. (ii) If a rally is won by the server, the latter serves in the next rally. Otherwise, the receiver becomes the server in the next rally. A point is scored after each rally. (iii) The winner of the *game* is the first player to score n points.

A *match* would typically consist of a sequence of such games, and the winner of the match is the first player to win M games. Actually, it is usually so that in game $m \geq 2$, the first server is not determined by flipping a coin, but rather according to some prespecified rule: the most common one states that the first server in game m is the winner in game $m - 1$, but alternatively, the players might simply take turns as the first server in each game until the match is over. It turns out that, in the probabilistic model we consider below, the probability that a fixed player wins the match is the same under both rules; see [1]. This clearly allows us to focus on a single game in the sequel—as in most previous works in the field (references will be given below). Extensions of our results to the match level can then trivially be obtained by appropriate conditioning arguments, taking into account the very rule adopted for determining the first server in each game.

The side-out scoring system has been used in various sports, sometimes up to tiny unimportant refinements, involving typically, in case of a tie at $n - 1$, the possibility (for the receiver) to choose whether the game should be played to $n + \ell$ (for some fixed $\ell \geq 2$) or to n ; see Section 2. When based on the so-called English scoring system, Squash currently uses $(n, M) = (9, 3)$. Racquetball is essentially characterized by $(n, M) = (15, 2)$ (the possible third game is actually played to 11 only). Until 2006, Badminton was using $(n, M) = (15, 2)$ and $(n, M) = (11, 2)$ for men's and women's singles, respectively—with an exception in 2002, where $(n, M) = (7, 3)$ was experimented. Volleyball, for which the term *persons* above should of course be understood as *teams*, was based on $(n, M) = (15, 3)$ until 2000. In both badminton and volleyball, this scoring system was then replaced with the rally-point system. Similarly, squash, at the international level, now is based on the American version of its scoring system, which is nothing but the rally-point system, in this case with $(n, M) = (11, 3)$. Investigating the deep implications of this transition from the original side-out scoring system to the rally-point scoring system was one of the main motivations for this work; see Section 4.

Irrespective of the scoring system adopted, the most common probabilistic model for the sequence of rallies assumes that the rally outcomes are i.i.d., in

the sense that they (i) are mutually independent (the probability that a player wins a rally is not affected by outcomes of the other rallies) and (ii) are, conditionally on the server, identically distributed (the probability that a player wins a rally when serving is constant over time). This implies that the game is governed by the parameter $(p_a, p_b) \in [0, 1] \times [0, 1]$, where the *rally-winning probability* p_a (respectively, p_b) is the probability that Player *A* (respectively, Player *B*) wins a rally when serving.

This model is the most widely accepted choice for mathematical analysis of sports like badminton (see references below), tennis (see [2, 3]) or table tennis (see [4]), although the existence of such player-related “governing parameters” may be disputable—discussion of this, and the consequences of using different modeling assumptions can be found in [5]. We will throughout refer to the above probabilistic model as the *server model*, in contrast with the *no-server model* in which any rally is won by *A* with probability p irrespective of the server, that is, the submodel obtained when taking $p = p_a = 1 - p_b$.

The probabilistic properties of a single game played under the side-out scoring system have been investigated in various works. Hsi and Burich [6] attempted to derive the probability distribution of game scores—in the sequel, we simply speak of the *score distribution*—in terms of p_a and p_b , but their derivation based on standard combinatorial arguments was wrong. The correct score distribution (hence also the resulting game-winning probabilities) was first obtained in [7] by applying results on sums of random variables having the modified geometric distribution. Keller [8] computed probabilities of very extreme scores, whereas Marcus [9] derived the complete score distribution in the no-server model. Strauss and Arnold [10], by identifying the point earning process as a Markov chain, obtained more directly the same general result as in [7]. They further used the score distribution to define maximum likelihood estimators and moment estimators of the rally-winning probabilities (both in the server and no-server models), and based on these estimates a ranking system (relying on Bradley–Terry paired comparison methods) for the players of a league or tournament. Simmons [11] determined the score distribution under the two scoring systems, this time by using a quick and direct combinatorial analysis of a single game. He discussed handicapping and strategies (for deciding whether the receiver should go for a game played to $n + \ell$ or not in case of a tie at $n - 1$), and attempted a comparison of the two scoring systems. More recently, Percy [12] used Monte Carlo simulations to compare game-winning probabilities and *expected* durations for both scoring systems in the no-server model.

To sum up, the score distributions have been obtained through several different probabilistic methods, and were used to discuss several aspects of the game. In contrast, the distribution of the number of rallies needed to complete a single game (D , say) remains virtually unexplored for the side-out scoring system (for the rally-point scoring system, the distribution of D is

simply determined by the score distribution). To the best of our knowledge, the only theoretical result on D under the side-out scoring system provides lower and upper bounds for the *expected value* of D ; see (20) in [11], or (2) below. Beyond the lack of exact results on D (only approximate theoretical results or simulation-based results are available so far), it should be noted that only the expected value of D has been studied in the literature. This is all the more surprising because, in various sports (e.g., in badminton and volleyball), uncertainty about D —which is related to its *variance*, not to its expected value—was one of the most important arguments to switch from the side-out scoring system to the rally-point scoring system. Exact results on the moments of D —or even better, its distribution—are then much desirable as they would allow to investigate whether the transition to the rally-point system indeed reduced uncertainty about D . More generally, precise results on the distribution of D would allow for a much deeper comparison of both scoring systems. They would also be of high practical relevance, e.g., to tournament organizers, who need planning their events and deciding in advance the number of matches—hence the number of players—the events will be able to host.

For the side-out scoring system, however, results on the distribution of D cannot be obtained from a *point-level* analysis of the game. That is the reason why the present work rather relies on a *rally-level* combinatorial analysis. This allows to get rid of the uncertainty about the number of rallies needed to score a single point, and results into an exact computation of the distribution of D —and actually, even of the number of rallies needed to achieve any fixed score. We derive explicitly the expectation and variance of D , and use our results to compare the two scoring systems not only in terms of game-winning probabilities, but also in terms of durations. Some of our findings are quite surprising, and unattainable through Monte Carlo experiments; see Section 5.

Our results reveal significant differences between both scoring systems, and help to explain why the transition from one scoring system to the other has more important implications than those predicted from game-winning probabilities alone. As suggested above, they could be used by tournament organizers to plan accurately their events, but also by national or international federations to better perform the possible transition from the side-out scoring system to the rally-point one; see Section 6 for a discussion. Also, our results allow for estimating, somewhat in the spirit of [2], the probability that a particular player wins a match (not only at the beginning, but at any stage during its progress), as well as forecasting the duration of the said match. This, of course, would have important applications for TV broadcast programmers, among others. Finally, our results open the way to efficient estimation of the rally-winning probabilities, based on observed scores and durations; see Section 6 for a discussion.

The outline of the paper is as follows. In Section 2, we describe our rally-level analysis of a single game played under the side-out scoring system,

and show that it also leads to the score distribution already derived in [7, 10], and [11]. Section 3 explains how this rally-level analysis further provides (i) the expectation and variance of the number of rallies needed to achieve a fixed score (Section 3.1) and also (ii) the corresponding exact distribution (Section 3.2). In Section 4, we then use our results to compare the side-out and rally-point scoring systems, both in terms of game-winning probabilities (Section 4.1) and durations (Section 4.2). In Section 5, we perform Monte Carlo simulations and compare the results with our theoretical findings. Section 6 presents the conclusion and provides some final comments. Finally, an appendix collects proofs of technical results.

2. Rally-level derivation of the score distribution under the side-out scoring system

In this section, we conduct our rally-level analysis of a single game played under the side-out scoring system. We will make the distinction between A -games and B -games, with the former (respectively, the latter) being defined as games in which Player A (respectively, Player B) is the first server. Wherever possible, we will state our results/definitions in the context of A -games only; in such cases the corresponding results/definitions for B -games can then be obtained by exchanging the roles played by A and B , that is, by exchanging (i) p_a and p_b and (ii) the number of points scored by each player. Whenever not specified, the server S will be considered random, and we will denote by $s_a := P[S = A]$ and $s_b := P[S = B] = 1 - s_a$ the probabilities that the game considered is an A -game and a B -game, respectively. This both covers games where the first server is determined by flipping a coin and games where the first server is fixed (by letting $s_a \in \{0, 1\}$).

Our rally-level analysis of the game will be based on the concepts of *interruptions* and *exchanges* first introduced in [6]. More precisely, we adopt

DEFINITION 1. *An A-interruption is a sequence of rallies in which B gains the right to serve from A, scores at least one point, then (unless the game is over) relinquishes the service back to A, who will score at least one point. An exchange is a sequence of two rallies in which one player gains the right to serve, but immediately loses this right before he/she scores any point (so that the potential of consecutive scoring by his/her opponent is not interrupted).*

We point out that A -interruptions are characterized in terms of score changes only (and in particular may contain one or several exchanges) and that, at any time, an exchange clearly occurs with probability $q := q_a q_b := (1 - p_a)(1 - p_b)$.

Now, for $C \in \{A, B\}$, denote by $E^{\alpha, \beta, C}(r, j)$ the event associated with a sequence of rallies that (i) gives rise to α points scored by Player A and β points scored by Player B, (ii) involves exactly r A -interruptions and j exchanges,

and (iii) is such that Player C scores a point in the last rally; the superscript C therefore indicates who is scoring the last point, and it is assumed here that $\alpha > 0$ (respectively, $\beta > 0$) if $C = A$ (respectively, if $C = B$). We will write

$$p_{C_1}^{\alpha, \beta, C_2}(r, j) := \mathbb{P}[E^{\alpha, \beta, C_2}(r, j) \mid S = C_1], \quad C_1, C_2 \in \{A, B\}.$$

We then have the following result (see the Appendix for the proof).

LEMMA 1. *Let $\gamma_0 := \min\{\beta, 1\}$, $\gamma_1 := \min\{\alpha, \beta\}$, and $\gamma_2 := \min\{\alpha, \beta - 1\}$. Then, setting $\binom{-1}{-1} := 1$, we have $p_A^{\alpha, \beta, A}(r, j) = \binom{\alpha + \beta + j - 1}{j} \binom{\alpha}{r} \binom{\beta - 1}{r - 1} p_a^\alpha \times p_b^\beta q^{r+j}$, $r \in \{\gamma_0, \dots, \gamma_1\}$, $j \in \mathbb{N}$, and $p_A^{\alpha, \beta, B}(r, j) = \binom{\alpha + \beta + j - 1}{j} \binom{\alpha}{r - 1} \binom{\beta - 1}{r - 1} p_a^\alpha \times p_b^\beta q a q^{r+j-1}$, $r \in \{1, \dots, \gamma_2 + 1\}$, $j \in \mathbb{N}$.*

By taking into account all possible values for the numbers of A -interruptions and exchanges, Lemma 1 quite easily leads to the following result (see the Appendix for the proof), which then trivially provides the score distribution in an A -game, hence also the corresponding game-winning probabilities.

THEOREM 1. *Let $p_{C_1}^{\alpha, \beta, C_2} := \mathbb{P}[E^{\alpha, \beta, C_2} \mid S = C_1]$, where $E^{\alpha, \beta, C_2} := \cup_{r, j} E^{\alpha, \beta, C_2}(r, j)$, with $C_1, C_2 \in \{A, B\}$. Then $p_A^{\alpha, \beta, A} = \frac{p_a^\alpha p_b^\beta}{(1-q)^{\alpha+\beta}} \sum_{r=\gamma_0}^{\gamma_1} \binom{\alpha}{r} \binom{\beta-1}{r-1} q^r$ and $p_A^{\alpha, \beta, B} = \frac{p_a^\alpha p_b^\beta q a}{(1-q)^{\alpha+\beta}} \sum_{r=1}^{\gamma_2+1} \binom{\alpha}{r-1} \binom{\beta-1}{r-1} q^{r-1}$.*

In the sequel, we denote game scores by couples of integers, where the first entry (respectively, second entry) stands for the number of points scored by Player A (respectively, by Player B). With this notation, a C -game ends on the score (n, k) (respectively, (k, n)), $k \in \{0, 1, \dots, n - 1\}$, with probability $p_C^{n, k, A}$ (respectively, $p_C^{k, n, B}$), hence is won by A (respectively, by B) with the (game-winning) probability

$$p_C^A := \mathbb{P}[E^A \mid S = C] = \sum_{k=0}^{n-1} p_C^{n, k, A}$$

(respectively, $p_C^B := 1 - p_C^A$); throughout, $E^A := \cup_{k=0}^{n-1} E^{n, k, A}$ (respectively, $E^B := \cup_{k=0}^{n-1} E^{k, n, B}$) denotes the event that the game—irrespective of the initial server—is won by A (respectively, by B). Of course, unconditional on the initial server, we have

$$\begin{aligned} p^{n, k, A} &:= \mathbb{P}[E^{n, k, A}] = p_A^{n, k, A} s_a + p_B^{n, k, A} s_b, \\ p^{k, n, B} &:= \mathbb{P}[E^{k, n, B}] = p_A^{k, n, B} s_a + p_B^{k, n, B} s_b, \end{aligned}$$

and

$$p^C := \mathbb{P}[E^C] = p_A^C s_a + p_B^C s_b,$$

for $C \in \{A, B\}$.

Figures 1(a)–(b) present, for an A -game with $n = 15$, the score distributions associated with $(p_a, p_b) = (0.7, 0.5), (0.6, 0.5), (0.5, 0.5)$, and $(0.4, 0.5)$. We reversed the k -axis in Figure 1(b), because, among all scores associated with a victory of B , the score $(14, 15)$ can be considered the closest to the score $(15, 14)$ (associated with a victory of A). It then makes sense to regard Figures 1(a)–(b) as a single plot. The resulting “global” probability curves are quite smooth and, as expected, unimodal (with the exception of the $p_a = p_b = 0.5$ curve, which is slightly bimodal). It appears that these score distributions are extremely sensitive to (p_a, p_b) , as are the corresponding game-winning probabilities (p_A^A ranges from 0.94 to 0.22, when, for fixed $p_b = 0.5$, p_a goes from 0.7 to 0.4). For $p_a = p_b = 0.5$, we would expect the global probability curve to be symmetric. The advantage Player A is given by serving first in the game, however, makes this curve slightly asymmetric; this is quantified by the corresponding probability that A wins the game, namely $p_A^A = 0.53 > 0.47 = p_A^B$.

As mentioned in the Introduction, sports based on the side-out scoring system may involve tiebreaks in case of a tie at $n - 1$. This means that, at this tie, the receiver has the option of playing through to n or “setting to ℓ ” (for a fixed $\ell \geq 2$), in which case the winner is the first player to score ℓ further points. For instance, games in the current side-out scoring system for squash are played to $n = 9$ points, and the receiver, at $(8, 8)$, may decide whether the game is to 9 or 10 points ($\ell = 2$). Before the transition to the rally-point system in 2006, similar tiebreak rules were used in badminton, there with $n = 15$ and $\ell = 3$. Assuming that the game is always set to ℓ in case of a tie at $n - 1$, the resulting score distribution can then be easily derived from Theorem 1 by appropriate conditioning; for instance, the score $(n + \ell - 1, n + k - 1)$, $k \in \{0, 1, \dots, \ell - 1\}$ occurs in an A -game with probability $p_A^{n-1, n-1, A} p_A^{\ell, k, A} + p_A^{n-1, n-1, B} p_B^{\ell, k, A}$. We stress that all results we derive in the later sections can also be extended to scoring systems involving tiebreaks, again by appropriate conditioning. Finally, various papers discuss tiebreak strategies (whether to play through or to set the game to ℓ) on the basis of p_a and p_b ; see, e.g., [11–14].

3. Distribution of the number of rallies under the side-out scoring system

As mentioned in the Introduction, the literature contains few results about the number of rallies D needed to complete a single game played under the side-out scoring system. Of course, the distribution of D can always be investigated by simulations; see, e.g., [12], where Monte Carlo methods are used to estimate the expectation of D for a broad range of rally-winning probabilities in the no-server model. To the best of our knowledge, the only available *theoretical* result is due to Simmons [11], and provides lower and upper bounds on the *expectation* of D in an A -game conditional on a victory of A on the score (n, k) . More specifically, letting

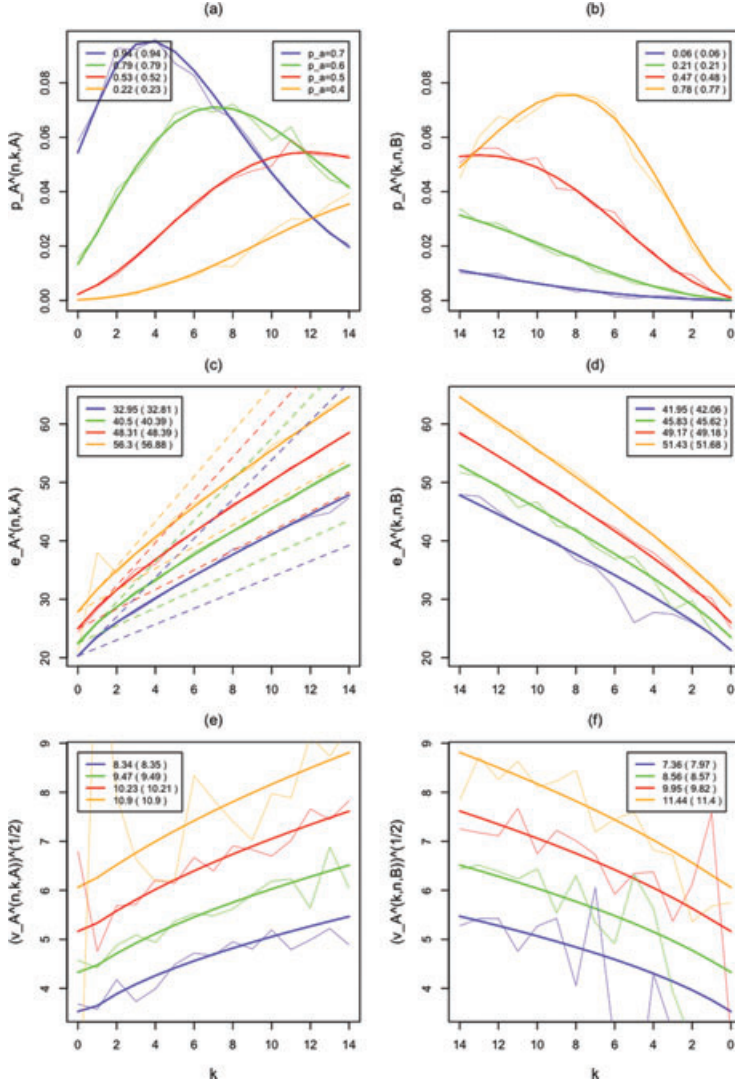


Figure 1. All subfigures refer to an A -game played under the side-out scoring system with $n = 15$. Left: for $(p_a, p_b) = (0.7, 0.5)$, $(0.6, 0.5)$, $(0.5, 0.5)$, and $(0.4, 0.5)$, (a) probabilities $p_A^{n,k,A}$ that Player A wins the game (score (n, k)) (along with the probabilities p_A^A that Player A wins the game), (c) expected values $e_A^{n,k,A}$, and (e) standard deviations $(v_A^{n,k,A})^{1/2}$ of the numbers of rallies D conditional on the corresponding events (along with the expected values e_A^A and standard deviations $(v_A^A)^{1/2}$ of D conditional on a victory of A). Right: the corresponding values for victories of B on the score (k, n) . As for the expected values and standard deviations of D unconditional on the score or the winner, we have $(e_A, v_A^{1/2}) = (33.5, 8.6)$, $(41.6, 9.5)$, $(48.7, 10.1)$, and $(52.5, 11.5)$, for $(p_a, p_b) = (0.7, 0.5)$, $(0.6, 0.5)$, $(0.5, 0.5)$, and $(0.4, 0.5)$, respectively. Estimated probabilities, expectations, and standard deviations based on 5,000 replications are also reported (thinner lines in plots and numbers between parentheses in legend boxes). Dashed lines in (c) correspond to the lower and upper bounds in (2); see [11].

$$e_{C_1}^{\alpha, \beta, C_2} := E[D \mid E^{\alpha, \beta, C_2}, S = C_1], \quad C_1, C_2 \in \{A, B\}, \quad (1)$$

Simmons' result states that

$$(n+k) \frac{1+q}{1-q} \leq e_A^{n, k, A} \leq (n+k) \frac{1+q}{1-q} + 2k, \quad k = 0, 1, \dots, n-1. \quad (2)$$

Unless a shutout is considered (that is, $k = 0$), this is only an approximate result, whose accuracy quickly decreases with k . Again, the reason why no exact results are available is that all analyses of the game in the literature are of a *point-level* nature. In sharp contrast, our *rally-level* analysis allows, *inter alia*, for obtaining exact values of all moments of D , as well as its complete distribution.

3.1. Moments

We first introduce the following notation. Let $R_A^{\alpha, \beta, A}$ (respectively, $R_A^{\alpha, \beta, B}$) be a random variable assuming values $r = \gamma_0, \gamma_0 + 1, \dots, \gamma_1$ (respectively, $r = 1, 2, \dots, \gamma_2 + 1$) with corresponding probabilities $W_A^{\alpha, \beta, A}(q, r) := \binom{\alpha}{r} \binom{\beta-1}{r-1} q^r / [\sum_{s=\gamma_0}^{\gamma_1} \binom{\alpha}{s} \binom{\beta-1}{s-1} q^s]$ (respectively, $W_A^{\alpha, \beta, B}(q, r) := \binom{\alpha}{r-1} \binom{\beta-1}{r-1} q^{r-1} / [\sum_{s=1}^{\gamma_2+1} \binom{\alpha}{s-1} \binom{\beta-1}{s-1} q^{s-1}]$). Conditioning with respect to the number of A -interruptions and exchanges then yields the following result (see the Appendix for the proof).

THEOREM 2. Let $t \mapsto M_{C_1}^{\alpha, \beta, C_2}(t) = E[e^{tD} \mid E^{\alpha, \beta, C_2}, S = C_1]$, $C_1, C_2 \in \{A, B\}$, be the moment generating function of D conditional on the event $E^{\alpha, \beta, C_2} \cap [S = C_1]$, and let $\delta_{C_1, C_2} = 1$ if $C_1 = C_2$ and 0 otherwise. Then

$$M_A^{\alpha, \beta, C}(t) = \left(\frac{(1-q)e^t}{1-qe^{2t}} \right)^{\alpha+\beta} E[e^{t(2R_A^{\alpha, \beta, C} - \delta_{B, C})}],$$

for $C \in \{A, B\}$.

Quite remarkably, those moment generating functions (hence also all resulting moments) depend on (p_a, p_b) through $q = (1-p_a)(1-p_b)$ only. Taking first and second derivatives with respect to t in the above expressions and setting $t = 0$ then directly yields the following closed form expressions for the expected values $e_{C_1}^{\alpha, \beta, C_2}$ from (1) and for the corresponding variances

$$v_{C_1}^{\alpha, \beta, C_2} := \text{Var}[D \mid E^{\alpha, \beta, C_2}, S = C_1], \quad C_1, C_2 \in \{A, B\}.$$

COROLLARY 1. For $C \in \{A, B\}$, we have (i) $e_A^{\alpha, \beta, C} = (\alpha + \beta) \frac{1+q}{1-q} - \delta_{B, C} + 2E[R_A^{\alpha, \beta, C}]$ and (ii) $v_A^{\alpha, \beta, C} = 4(\alpha + \beta) \frac{q}{(1-q)^2} + 4\text{Var}[R_A^{\alpha, \beta, C}]$. Moreover, (iii) $e_A^{\alpha, \beta, C}$ is strictly monotone increasing in q .

Clearly, Corollary 1 confirms Simmons' result that the expected number of rallies in an A -game won by A on the score (n, k) is $e_A^{n,k,A} = n \frac{1+q}{1-q}$ for $k = 0$. More interestingly, it also shows that the exact value for any $k > 0$ is given by

$$e_A^{n,k,A} = (n+k) \frac{1+q}{1-q} + 2 \sum_{r=1}^k r W_A^{n,k,A}(q, r), \quad k = 1, \dots, n-1. \quad (3)$$

Note that this is compatible with Simmons' result in (2) because the second term in the right-hand side of (3) is a weighted mean of $2r$, $r = 1, \dots, k$. Similarly, the expected number of rallies in an A -game won by B on the score (k, n) , $k = 0, 1, \dots, n-1$, is $e_A^{k,n,B} = (n+k) \frac{1+q}{1-q} - 1 + 2 \sum_{r=1}^{k+1} r W_A^{k,n,B}(q, r)$.

The expectation and variance of D , in a C -game won by A , are then given by

$$\begin{cases} e_C^A := E[D | E^A, S = C] = \frac{1}{p_C^A} \sum_{k=0}^{n-1} p_C^{n,k,A} e_C^{n,k,A} \\ v_C^A := \text{Var}[D | E^A, S = C] = \left[\frac{1}{p_C^A} \sum_{k=0}^{n-1} p_C^{n,k,A} \left(v_C^{n,k,A} + (e_C^{n,k,A})^2 \right) \right] - (e_C^A)^2, \end{cases} \quad (4)$$

while, in a C -game unconditional on the winner, they are given by

$$\begin{cases} e_C := E[D | S = C] = p_C^A e_C^A + p_C^B e_C^B, \\ v_C := \text{Var}[D | S = C] = (v_C^A + (e_C^A)^2) p_C^A + (v_C^B + (e_C^B)^2) p_C^B - (e_C)^2. \end{cases} \quad (5)$$

Finally, unconditional on the server, this yields

$$\begin{cases} e^A := E[D | E^A] = e_A^A s_a + e_B^A s_b, & e := E[D] = e_A s_a + e_B s_b, \\ v^A := \text{Var}[D | E^A] = (v_A^A + (e_A^A)^2) s_a + (v_B^A + (e_B^A)^2) s_b - (e^A)^2, \\ v := \text{Var}[D] = (v_A + e_A^2) s_a + (v_B + e_B^2) s_b - e^2. \end{cases} \quad (6)$$

Figures 1(c)–(f) plot, for $n = 15$, $e_A^{n,k,A}$, $e_A^{k,n,B}$, $(v_A^{n,k,A})^{1/2}$, and $(v_A^{k,n,B})^{1/2}$ versus k for $(p_a, p_b) = (0.7, 0.5)$, $(0.6, 0.5)$, $(0.5, 0.5)$, and $(0.4, 0.5)$, and report the corresponding numerical values of e_A^A , e_A^B , e_A , $(v_A^A)^{1/2}$, $(v_A^B)^{1/2}$, and $(v_A)^{1/2}$. All expectation and standard deviation curves appear to be strictly monotone increasing functions of the number $(n+k)$ of points scored, which was maybe expected. More surprising is the fact that—if one discards very small values of k —these curves are also roughly linear. Clearly, Simmons' lower and upper bounds (2), which are plotted versus k in Figure 1(c), only provide poor approximations of the exact expected values, particularly so for large k .

The dependence on (p_a, p_b) may be more interesting than that on k . Note that, for each k , $e_A^{n,k,A}$ and $e_A^{k,n,B}$ (hence also, e_A^A , e_A^B , and e_A) are decreasing functions of p_a , which confirms Corollary 1(iii). Similarly, all quantities related to standard deviations also seem to be decreasing functions of p_a . Now, it is

seen that, as a function of p_a , the expectation e_A^A is more spread out than e_A^B . Indeed, the former ranges from 32.95 ($p_a = 0.7$) to 56.30 ($p_a = 0.4$), whereas the latter ranges from 41.95 to 51.43. On the contrary, the standard deviation of D is more concentrated in an A -game won by A (where it ranges from 8.34 ($p_a = 0.7$) to 10.90 ($p_a = 0.4$)) than in an A -game won by B (where it ranges from 7.36 to 11.44). This phenomenon will appear even more clearly in Figure 3 below, where the same values of (p_a, p_b) are considered. Note that the values of e_A^A , e_A^B , and e_A are totally in line with the score distribution and the expected values of D for each scores. For instance, the value $e_A^B = 41.95$ for $p_a = 0.7$ translates the fact that when B wins such an A -game, it is very likely (see Figure 1(b)) that he/she will do so on a score that is quite tight, resulting on a large expected value for D (whereas, *a priori*, the values of $e_A^{k,n,B}$ range from 47.82 to 21.29 when k goes from 14 to 0). The dependence of the expectation and standard deviation of D on rally-winning probabilities will further be investigated in Section 4 for the no-server model when comparing the side-out scoring system with its rally-point counterpart.

Finally, in the case $p_a = p_b = 0.5$, the fact that A is the first server in the game again brings some asymmetry in the expected values and standard deviations of D ; in particular, this serve advantage alone is responsible for the fact that $48.31 = e_A^A < e_A^B = 49.17$, and, maybe more mysteriously, that $10.23 = (v_A^A)^{1/2} > (v_A^B)^{1/2} = 9.95$.

3.2. Distribution

The moment generating functions given in Theorem 2 allow, through a suitable change of variables, for obtaining the corresponding probability generating functions. These can in turn be rewritten as power series whose coefficients yield the distribution of D conditional on the event $E^{\alpha,\beta,C} \cap [S = A]$ (see the Appendix for the proof).

THEOREM 3. *Let $z \mapsto G_{C_1}^{\alpha,\beta,C_2}(z) = E[z^D | E^{\alpha,\beta,C_2}, S = C_1]$, $C_1, C_2 \in \{A, B\}$, be the probability generating function of D conditional on the event $E^{\alpha,\beta,C_2} \cap [S = C_1]$. Then, for $C \in \{A, B\}$,*

$$G_A^{\alpha,\beta,C}(z) = \frac{p_a^\alpha p_b^\beta q_a^{\delta_{B,C}}}{p_A^{\alpha,\beta,C}} \sum_{j=0}^{\infty} q^j H_A^{\alpha,\beta,C}(j) z^{\alpha+\beta+2j+\delta_{B,C}},$$

where, writing $m^+ := \max(m, 0)$, we let

$$H_A^{\alpha,\beta,A}(j) := \sum_{l=(j-\gamma_1)^+}^j \binom{\alpha+\beta+l-1}{l} \binom{\alpha}{j-l} \binom{\beta-1}{j-l-1}$$

and

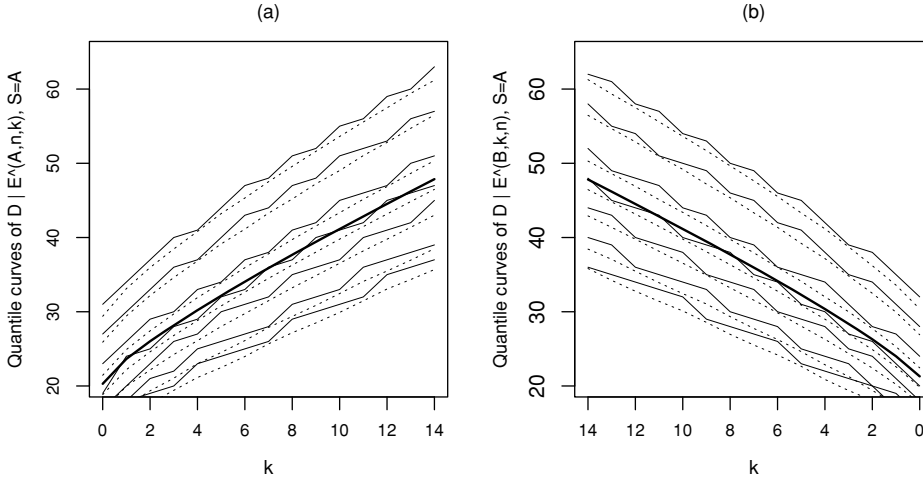


Figure 2. Both subfigures refer to an A -game played under the side-out scoring system with $n = 15$ and $(p_a, p_b) = (0.6, 0.5)$. Subfigure (a) (respectively, Subfigure (b)) reports, as a function of k , the α -quantile of the number of rallies needed to complete the game, conditional on a victory of A on the score (n, k) (respectively, conditional on a victory of B on the score (k, n)), with $\alpha = 0.01, 0.05, 0.25, 0.50, 0.75, 0.95$, and 0.99 . Solid lines (respectively, dotted lines) correspond to standard (respectively, interpolated) quantiles; see Section 3.2 for details. The thicker solid curves give the expected values of D conditional on the same events, hence are the same as in Figure 1(c)–(d).

$$H_A^{\alpha, \beta, B}(j) := \sum_{l=(j-\gamma_2)^+}^j \binom{\alpha + \beta + l - 1}{l} \binom{\alpha}{j-l} \binom{\beta - 1}{j-l}.$$

This result gives the probability distribution of D , conditional on $E^{\alpha, \beta, C} \cap [S = A]$, for $C \in \{A, B\}$. Note that, as expected, we have $P[D = d | E^{\alpha, \beta, A}, S = A] = 0 = P[D = d + 1 | E^{\alpha, \beta, B}, S = A]$ for all $d < \alpha + \beta$. Moreover, for all nonnegative integer j , $P[D = \alpha + \beta + 2j + 1 | E^{\alpha, \beta, A}, S = A] = 0 = P[D = \alpha + \beta + 2j | E^{\alpha, \beta, B}, S = A]$. In the sequel, we refer to this as the *server-effect*.

Theorem 3 of course allows for investigating the shape of the distribution of D above all scores, and not only, as in Figures 1(c)–(f), its expectation and standard deviation. This is what is done in Figure 2, which plots, as a function of the score, quantiles of order $\alpha = 0.01, 0.05, 0.25, 0.5, 0.75, 0.95$, and 0.99 for $(p_a, p_b) = (0.6, 0.5)$. For each α , two types of quantiles are reported, namely (i) the standard quantile $q_\alpha := \inf\{d : P[D \leq d | E^{\alpha, \beta, C}, S = A] \geq \alpha\}$ and (ii) an interpolated quantile, for which the interpolation is conducted linearly over the set $(d, d + 2)$ containing the expected quantile (here, we avoid interpolating over $(d, d + 1)$ because of the above server-effect, which implies that either d or $d + 1$ does not bear any probability mass). One of the most prominent features of Figure 2 is the wiggleness of the standard quantile curves, which is directly associated with the server-effect. It should be noted

that the expectation curves (which are the same as in Figures 1(c)–(d)) stand slightly above the median curves, which possibly indicates that, above each score, the conditional distribution of D is somewhat asymmetric to the right. This (light) asymmetry is confirmed by the other quantiles curves.

Now, the probability distribution of D in an A -game, unconditional on the score, is of course derived trivially from its conditional version obtained above and the score distribution of Section 2. The general form of this distribution is somewhat obscure (and will not be explicitly given here), but it yields easily interpretable expressions for small values of d . For instance, one obtains

$$\begin{aligned} P[D = n \mid S = A] &= p_a^n, \\ P[D = n + 1 \mid S = A] &= q_a p_b^n, \\ P[D = n + 2 \mid S = A] &= n q p_a^n + p_a q_a p_b^n, \dots \end{aligned}$$

Finally, the unconditional distribution of D is simply obtained through $P[D = k] = P[D = k \mid S = A]s_a + P[D = k \mid S = B]s_b$, $k \geq 0$, where one computes the distribution for a B -game by inverting p_a and p_b in the distribution for an A -game.

Figure 3 shows that there are a number of remarkable aspects to these distributions. First note the influence of the above mentioned server-effect, which causes the wiggleness visible in most curves there. Also note that the distributions in Figure 3(c) are much less wiggling than the corresponding curves in Figures 3(a)–(b). As it turns out, this wiggleness is present, albeit more or less markedly, at all stages (that is, not only to the right of the mode) for every choice of (p_a, p_b) . Most importantly, despite their irregular aspect, all curves are essentially unimodal, as expected.

Now, consider the dependence on p_a of the position and spread of these curves. One sees that while their spread clearly increases much more rapidly with p_a in Figure 3(b) than in Figure 3(a), the opposite can be said for their mode. This is easily understood in view of the corresponding means and variances, which are recalled in the legend boxes (and coincide with those from Figure 1). As for the curves in Figure 3(c), they are obtained by averaging the corresponding curves in Figure 3(a) and Figure 3(b) with weights p_A^A and $p_A^B = 1 - p_A^A$, respectively. Taking into account the values of these probabilities explains why the curves with $p_a = 0.7$ and $p_a = 0.6$ are essentially the corresponding curves in Figure 3(a), whereas that with $p_a = 0.4$ is closer to the corresponding curve in Figure 3(b).

4. Comparison with the rally-point scoring system

One of the main motivations for this work was to compare more deeply the side-out scoring system considered in Sections 2 and 3 with the *rally-point*

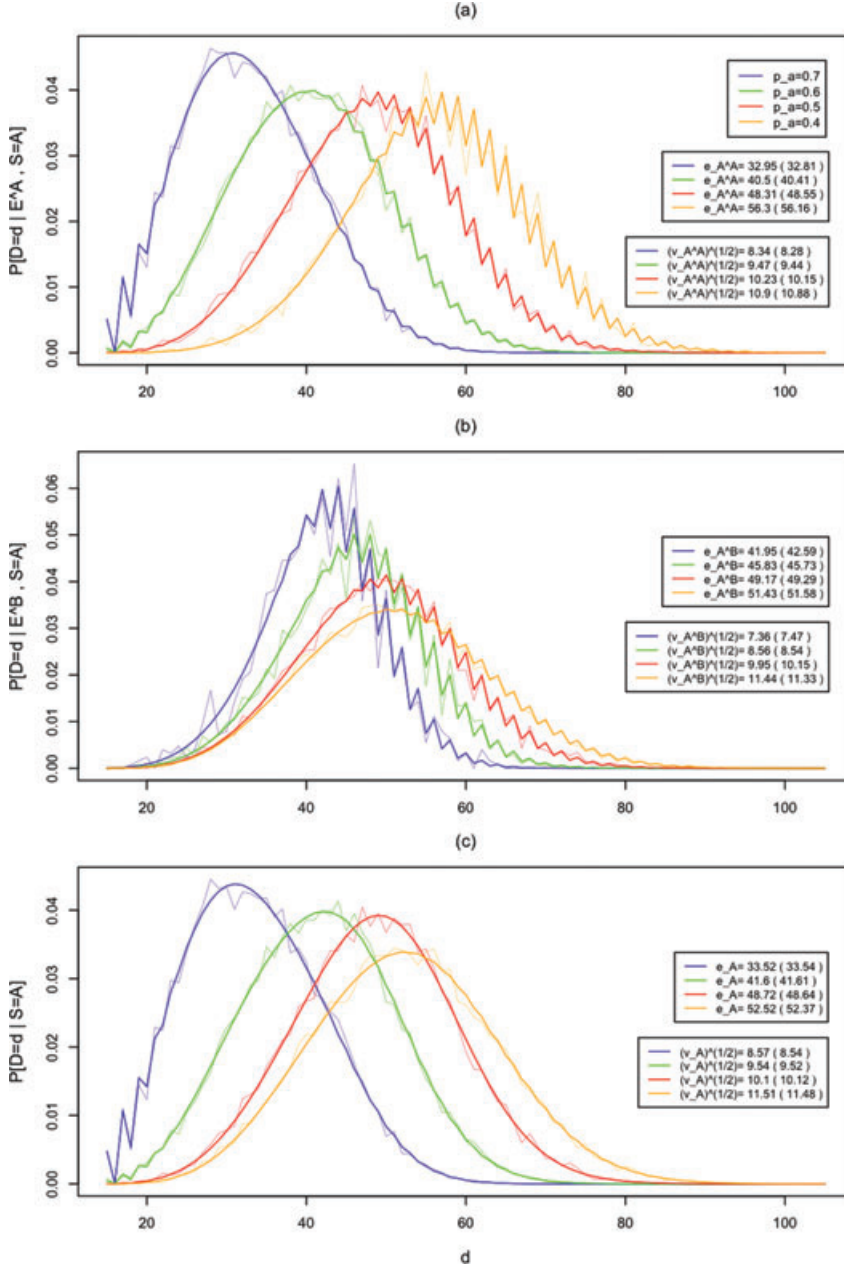


Figure 3. All subfigures refer to an A -game played under the side-out scoring system with $n = 15$. For $(p_a, p_b) = (0.7, 0.5)$, $(0.6, 0.5)$, $(0.5, 0.5)$, and $(0.4, 0.5)$, they report the probabilities that the number of rallies D needed to complete the game takes value d , (a) conditional upon a victory of Player A , (b) conditional upon a victory of Player B , and (c) unconditional. Estimated probabilities, expectations, and standard deviations based on 20,000 replications are also reported (thinner lines in plots and numbers between parentheses in legend boxes).

scoring system. As mentioned in the Introduction, many sports recently switched (e.g., badminton, volleyball)—or are in the process of switching (e.g., squash)—from the side-out scoring system to its rally-point counterpart, whereas others (e.g., racquetball) so far are sticking to the side-out scoring system. It is therefore natural to investigate the implications of the transition to the rally-point system.

The literature, however, has focused on the impact of the scoring system on the *outcome* of the game—studied by comparing the game-winning probabilities under both scoring systems; see, e.g., [11]. This is all the more surprising because there have been, in the sport community, much debate and questions about how much the duration of the game is affected by the scoring system. Moreover, it is usually reported that the main motivation for turning to the rally-point system is to regulate the playing time (that is, to make the length of the match more predictable), which is of primary importance for television, for instance. Whether the transition to the rally-point system has indeed served that goal, and, if it has, to what extent, are questions that have not been considered in the literature, and were at best addressed on empirical grounds only (by international sport federations).

In this section, we will provide an in-depth comparison of the two scoring systems, both in terms of game-winning probabilities and in terms of durations, which will provide theoretical answers to the questions above. Again, this is made possible by our rally-level analysis of the game and the results of the previous sections on the distribution of the number of rallies under the side-out scoring system. As we will discuss in Section 6, our results are potentially of high interest both for international federations and for local tournament organizers.

4.1. Game-winning probabilities

Although the game-winning probabilities for an A -game played under the rally-point system have already been obtained in the literature (see, e.g., [11]), we start by deriving them quickly, mainly for the sake of completeness, but also because they easily follow from the combinatorial methods used in the previous sections. First note that there cannot be *exchanges* in the rally-point scoring system, as it is understood in Definition 1 that no point is scored in an exchange. We then denote by $\bar{E}_A^{\alpha,\beta,C}(r)$ ($C \in \{A, B\}$) the event associated with a sequence of rallies that, in the rally-point system, (i) gives rise to α points scored by Player A and β points scored by Player B, (ii) involves exactly r A -interruptions, and (iii) is such that Player C scores a point in the last rally; again, it is assumed here that $\alpha > 0$ (respectively, $\beta > 0$) if $C = A$ (respectively, if $C = B$). We write

$$\bar{p}_{C_1}^{\alpha,\beta,C_2}(r) := \mathbb{P}[\bar{E}_A^{\alpha,\beta,C_2}(r) \mid S = C_1], \quad C_1, C_2 \in \{A, B\}.$$

The following result then follows along the same lines as for Lemma 1 and Theorem 1.

THEOREM 4. (i) *With the notation above, $\bar{p}_A^{\alpha,\beta,A}(r) = \binom{\alpha}{r} \binom{\beta-1}{r-1} p_a^{\alpha-r} p_b^{\beta-r} (q_a q_b)^r$, $r \in \{\gamma_0, \dots, \gamma_1\}$, and $\bar{p}_A^{\alpha,\beta,B}(r) = \binom{\alpha}{r-1} \binom{\beta-1}{r-1} p_a^{\alpha-r+1} p_b^{\beta-r} q_a (q_a q_b)^{r-1}$, $r \in \{1, \dots, \gamma_2 + 1\}$. (ii) *Writing $\bar{p}_A^{\alpha,\beta,C}$ for the probability of the event $\bar{E}_A^{\alpha,\beta,C} := \cup_r \bar{E}_A^{\alpha,\beta,C}(r)$, we have $\bar{p}_A^{\alpha,\beta,A} = p_a^\alpha p_b^\beta \sum_{r=\gamma_0}^{\gamma_1} \binom{\alpha}{r} \binom{\beta-1}{r-1} (t_a t_b)^r$ and $\bar{p}_A^{\alpha,\beta,B} = p_a^\alpha - p_b^{\beta-1} q_a \sum_{r=1}^{\gamma_2+1} \binom{\alpha}{r-1} \binom{\beta-1}{r-1} (t_a t_b)^{r-1}$, where we let $t_a = q_a/p_a$ and $t_b = q_b/p_b$.**

Remark 1. These expressions further simplify in the no-server model ($p := p_a = 1 - p_b$). There we indeed have $t_b = t_a^{-1}$, so that the above formulas yield $\bar{p}_A^{\alpha,\beta,A} = (\alpha + \beta - 1) p^\alpha (1 - p)^\beta$ and $\bar{p}_A^{\alpha,\beta,B} = (\alpha + \beta - 1) p^\alpha (1 - p)^\beta$.

Of course, the resulting score distribution and game-winning probabilities for an A -game directly follow from Theorem 4. In accordance with the notation adopted for the side-out scoring system, we will write

$$\begin{aligned} \bar{p}_C^A &:= \mathbb{P}[\bar{E}^A | S = C] := \mathbb{P}[\cup_{k=0}^{n-1} E^{n,k,A} | S = C] := \sum_{k=0}^{n-1} \bar{p}_C^{n,k,A}, \quad \bar{p}_C^B := 1 - \bar{p}_C^A, \\ \bar{p}^{n,k,A} &:= \mathbb{P}[\bar{E}^{n,k,A}] = \bar{p}_A^{n,k,A} s_a + \bar{p}_B^{n,k,A} s_b, \\ \bar{p}^{k,n,B} &:= \mathbb{P}[\bar{E}^{k,n,B}] = \bar{p}_A^{k,n,B} s_a + \bar{p}_B^{k,n,B} s_b, \end{aligned}$$

and

$$\bar{p}^C := \mathbb{P}[\bar{E}^C] = \bar{p}_A^C s_a + \bar{p}_B^C s_b.$$

Figures 4(a)–(b) plot the same score distribution curves as in Figures 1(a)–(b), respectively, but in the case of an A -game played under the rally-point scoring system with $n = 21$. Both pairs of plots look roughly similar, although extreme scores seem to be less likely in the rally-point scoring; this confirms the findings from [11] according to which shutouts are less frequent under the rally-point scoring system. Note also that, unlike for the side-out scoring, the $(p_a, p_b) = (0.5, 0.5)$ curve in Figure 4(a) is the exact reverse image of the corresponding one in Figure 4(b): for the rally-point scoring, Player A of course does not get any advantage from serving first if $(p_a, p_b) = (0.5, 0.5)$, which is confirmed by the game-winning probabilities $\bar{p}_A^A = \bar{p}_A^B = 0.5$.

Again, the dependence of the game-winning probabilities on (p_a, p_b) is of primary importance. We will investigate this dependence visually and compare it with the corresponding dependence for the side-out scoring system. To do so, we focus on the no-server version ($p = p_a = 1 - p_b$) of Badminton, where, as already mentioned, the side-out scoring system with $n = 15$ (men's

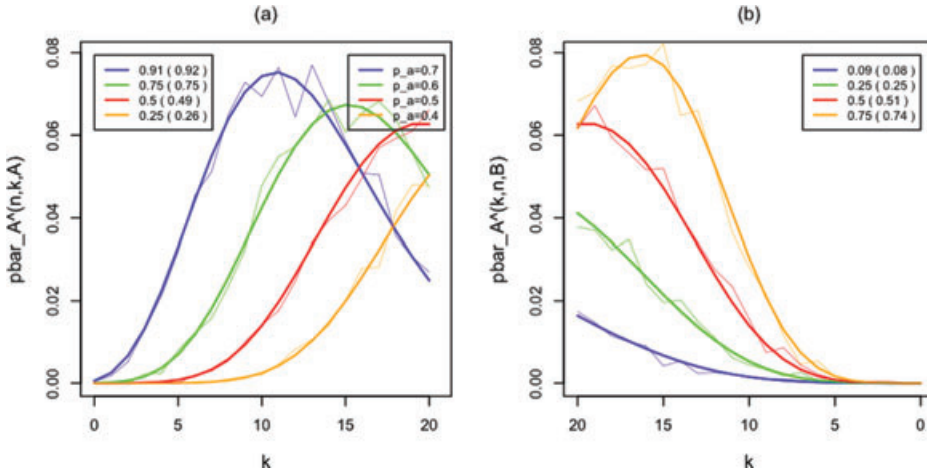


Figure 4. Both subfigures refer to an A -game played under the rally-point scoring system with $n = 21$. Subfigure (a): for $(p_a, p_b) = (0.7, 0.5)$, $(0.6, 0.5)$, $(0.5, 0.5)$, and $(0.4, 0.5)$, probabilities $\bar{p}_A^{n,k,A}$ that Player A wins the game on the score (n, k) , along with the probabilities \bar{p}_A^A that Player A wins the game. Subfigure (b): the corresponding values for victories of B on the score (k, n) . Estimated probabilities based on 5,000 replications are also reported (thinner lines in plots and numbers between parentheses in legend boxes).

singles) was recently replaced with the rally-point one characterized by $n = 21$. The results are reported in Figures 5(a)–(b). Figure 5(a) supports the claim—reported, e.g., in [11] or [12]—stating that, for any fixed p , the scoring barely influences game-winning probabilities. Now, while Figure 5(b) shows that the probability that Player A wins an A -game is essentially the same for both scoring systems if he/she is the best player ($\bar{p}_A^A/p_A^A \in (0.926, 1)$ for $p \geq 0.5$, and $\bar{p}_A^A/p_A^A \in (0.997, 1)$ for $p > 0.7$), it tells another story for $p < 0.5$: there, the probability that A wins an A -game played under the rally-point system (i) becomes relatively negligible for very small values of p (in the sense that $\bar{p}_A^A/p_A^A \rightarrow 0$ as $p \rightarrow 0$) and (ii) can be up to 28 times larger than under the side-out system (for values of p close to 0.1). Of course, one can say that (i) is irrelevant because it is associated with an event (namely, a victory of A) occurring with very small probability; (ii), however, constitutes an important difference between both scoring systems for values of p that are not so extreme.

4.2. Durations

In the rally-point system, the number of rallies needed to achieve the event $\bar{E}^{\alpha,\beta,C_2} \cap [S = C_1]$ is not random: with obvious notation, it is almost surely equal to $\bar{e}_{C_1}^{\alpha,\beta,C_2} = \alpha + \beta$, which explains why Figure 4 does not contain the rally-point counterparts of Figure 1(c)–(f). The various conditional and

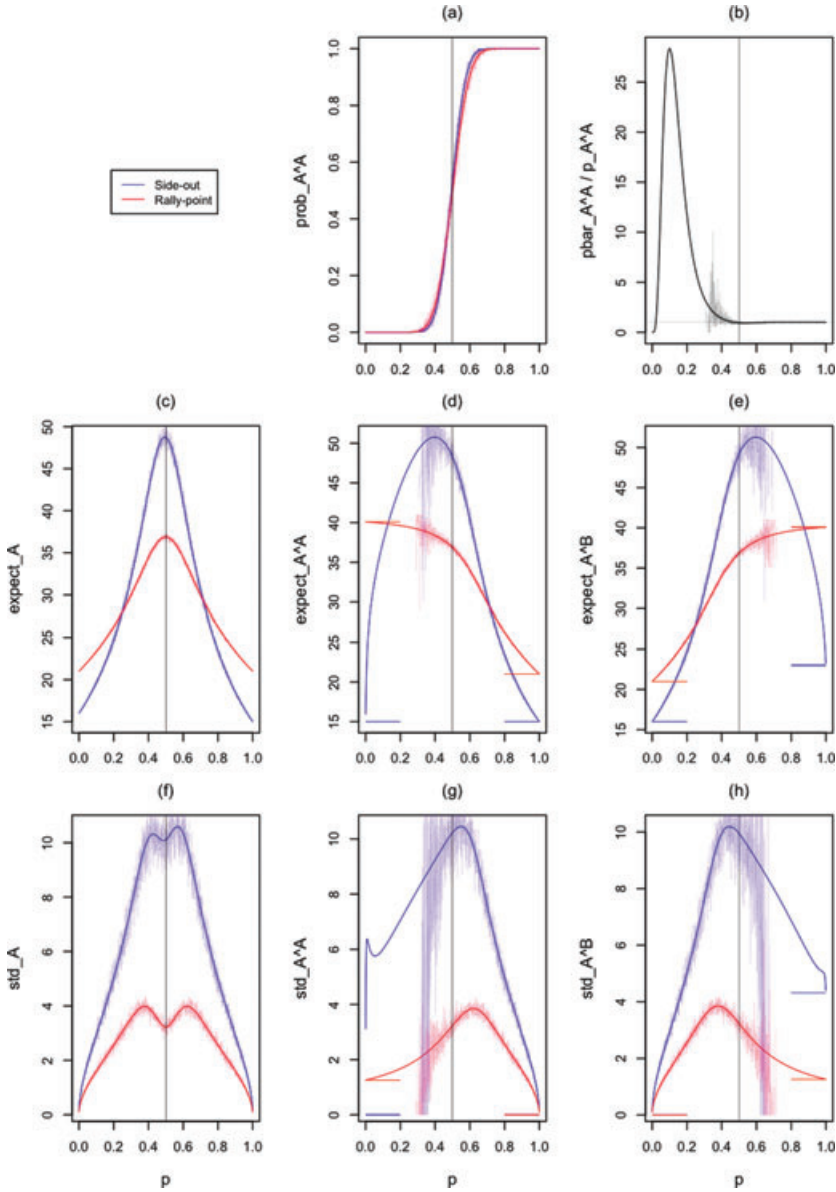


Figure 5. As a function of $p = p_a = 1 - p_b$ (hence, in the no-server model), probabilities $p_A^{15,k,A}$ (in blue) that Player A wins an $n = 15$ side-out A -game on the score $(15,k)$, along with the probabilities $\bar{p}_A^{21,k,A}$ (in red) that Player A wins an $n = 21$ rally-point A -game on the score $(21,k)$. Expectations (first row) and standard deviations (second row) of the number of rallies needed to complete the corresponding games, unconditional on the winner (first column), conditional on a victory of Player A (second column), and conditional on a victory of Player B (third column). Estimated probabilities, expectations, and standard deviations (based on 200 replications at each value of $p = 0, 0.0005, 0.0010, 0.0015, \dots, 0.9995$) are also reported (thinner lines).

unconditional means and variances of the number of rallies in the rally-point system (that is, the quantities \bar{e}_C^A , \bar{v}_C^A , \bar{e}_C , \bar{v}_C , \bar{e}^A , \bar{v}^A , \bar{e} , \bar{v}) can then be readily computed from the game-winning probabilities given in Theorem 4, in the exact same way as in (4)–(6) for the side-out scoring system. More generally, the corresponding distribution of the number of rallies in a game trivially follows from the same game-winning probabilities.

Figures 5(c)–(h) plot, as functions of $p = p_a = 1 - p_b$ (hence, in the no-server model), expected values and standard deviations of the numbers of rallies needed to complete (i) A -games played under the side-out system with $n = 15$ and (ii) A -games played under the rally-point system with $n = 21$. Clearly, those plots allow for an in-depth (original) comparison of both scoring systems. Let us first focus on durations unconditional on the winner of the game. Figure 5(c) shows that games played under the side-out system will last longer than those played under the rally-point one for players of roughly the same level (which was expected because the side-out system will then lead to many exchanges), whereas the opposite is true when one player is much stronger (which is explained by the fact that shutouts require more rallies in the rally-point scoring considered than in the side-out one). Maybe less expected is the fact (Figure 5(f)) that the standard deviation of D is, uniformly in $p \in (0, 1)$, smaller for the rally-point scoring system than for the side-out system, which shows that the transition to the rally-point system indeed makes the length of the game more predictable. The twin-peak shape of both standard deviation curves is even more surprising. Finally, note that, while the rally-point curves in Figures 5(c) and (f) are symmetric about $p = 0.5$, the side-out curves are not, which is due to the server-effect. This materializes into the limits of e_A given by 16 and 15 as $p \rightarrow 0$ and $p \rightarrow 1$, respectively (which was expected: if Player B wins each rally with probability one, he/she will indeed need 16 rallies to win an A -game, because he/she has to regain the right to serve before scoring his/her first point), but also translates into (i) the fact that the mode of the side-out curve in Figure 5(c) is not exactly located in $p = 0.5$ and (ii) the slightly different heights of the two local (side-out) maxima in Figure 5(f).

We then turn to durations conditional on the winner of the game, whose expected values and standard deviations are reported in Figures 5(d), (e), (g), and (h). These figures look most interesting and reveal important differences between both scoring systems. Even the general shape of the curves there are of a different nature for both scorings; for instance, the rally-point curves in Figures 5(d)–(e) are monotonic, while the side-out ones are unimodal. Similarly, in Figure 5(g), the rally-point curve is unimodal, whereas the side-out curve exhibits a most unexpected bimodal shape. It is also interesting to look at limits as $p \rightarrow 0$ or $p \rightarrow 1$ in those four subfigures; these limits, which are derived in Appendix A.3, are plotted as short horizontal lines. Consider first limits above events occurring with probability one, that is, limits as $p \rightarrow 1$ in Figures 5(d),

(g) and limits as $p \rightarrow 0$ in Figures 5(e), (h). The resulting limits are totally in line with our intuition: the four conditional standard deviations go to zero, which implies that the limiting conditional distribution of D simply is almost surely equal to the corresponding limiting (conditional) expectations. The latter themselves assume very natural values: for instance, for the same reason as above, e_A^B converges to 16, which is therefore the limit of D in probability.

Much more surprising is what happens for limits above events occurring with probability zero, that is, limits as $p \rightarrow 0$ in Figures 5(d), (g) and limits as $p \rightarrow 1$ in Figures 5(e), (h). Focusing first on the side-out scoring system, it is seen that a (miraculous) victory of A will require, in the limit, almost surely $D = 15$ points, while the limiting conditional distribution of D for victories of B is nondegenerate. The latter distribution is shown (see Appendix A.3) to be *uniform over* $\{n + 1, n + 2, \dots, 2n\}$ (hence is stochastically bounded!), which is compatible with the values $n + 1 + (n - 1)/2 (\approx 3n/2)$ and $(n - 1)^2/12$ for the limiting expectation and variance, respectively. It should be noted here that this huge difference between those two limiting conditional distributions of D is entirely due to the server-effect. In the absence of the server-effect, the Figures 5(e) and (h) should indeed be the exact reverse image of the Figures 5(d) and (g), respectively. Similarly, the bimodality of the side-out curve in Figure 5(g) is also due to the server-effect. We then consider the rally-point scoring, which is not affected by the server-effect, so that it is sufficient to consider at the limits as $p \rightarrow 0$ in Figures 5(d), (g). There, one also gets a nondegenerate limiting conditional distribution for D , with expectation $2n^2/(n + 1) (\approx 2n)$ and variance $2n^2(n - 1)/[(n + 1)^2(n + 2)] (\approx 2)$.

5. Simulations

We performed several Monte Carlo simulations, one for each figure considered so far (except Figure 2, as it already contains many theoretical curves). To describe the general procedure, we focus on the Monte Carlo experiment associated with the side-out scoring system in Figure 5 (results for the rally-point scoring system there or for the other figures are obtained similarly). For each of the 1,999 values of p considered in Figure 5, the corresponding values of $p_A^A(p)$, $e_A(p)$, $v_A(p)$, $e_A^C(p)$, $v_A^C(p)$, $C \in \{A, B\}$, were estimated on the basis of $J = 200$ independent replications of an A -game played under the side-out scoring system with $p_a = 1 - p_b = p$. Of course, for each fixed p , the game-winning probability $p_A^C(p)$ is simply estimated by the proportion of games won by C in the J corresponding A -games:

$$\hat{p}_A^C(p) := \frac{J^C}{J} := \frac{1}{J} \sum_{j=1}^J I_j^C,$$

where $I_j^C, j = 1, \dots, J$, is equal to one (respectively, zero) if Player C won (respectively, lost) the j th game. The corresponding estimates for $e_A(p), v_A(p), e_A^C(p)$, and $v_A^C(p)$ then are given by

$$\begin{aligned}\hat{e}_A(p) &:= \frac{1}{J} \sum_{j=1}^J d_j & \hat{v}_A(p) &:= \frac{1}{J} \sum_{j=1}^J (d_j - \hat{e}_A(p))^2, \\ \hat{e}_A^C(p) &:= \frac{1}{J^C} \sum_{j=1}^J d_j I_j^C, & \text{and } \hat{v}_A^C(p) &:= \frac{1}{J^C} \sum_{j=1}^J (d_j - \hat{e}_A^C(p))^2 I_j^C, \quad (7)\end{aligned}$$

where $d_j, j = 1, \dots, J$, is the total number of rallies in the j th game. These estimates are plotted in thin blue lines in Figure 5. Clearly, these simulations validate our theoretical results in Figures 5(a), (c), and (f). To describe what happens in the other plots, consider, e.g., Figure 5(g). There, it appears that the theoretical results are confirmed for large values of p only. However, this is simply explained by the fact that for small values of p , the denominator of $\hat{v}_A^A(p)$ (see (7)) is very small. Actually, among the 542×200 A -games associated with the 542 values of $p \leq 0.2710$, not a single one here led to a victory of A , so that the corresponding estimates $\hat{v}_A^A(p)$ are not even defined. Of course, values of p slightly larger than 0.2710 still give rise to a small number of victories of A , so that the corresponding estimates $\hat{v}_A^A(p)$ are highly unreliable. The situation improves substantially as p increases, as it can be seen in Figure 5(g). Figures 5(b), (d), (e), and (h) can be interpreted exactly in the same way.

This underlines the fact that expectations and variances conditional on events with small probabilities are extremely difficult—if not impossible—to estimate. To quantify this, let us focus again on Figure 5(g), and consider the local maximum on the left of the plot, which is (on the grid of values of p at hand) located in $p_0 := 0.0085$. The probability $p_A^A(p_0)$ of a victory of A in an A -game played under the side-out scoring system with $p = p_0$ is about 3.5×10^{-31} . Estimating $v_A^A(p_0)$ with the same accuracy as that achieved for, e.g., $v_A^A(0.5)$ in Figure 5(g) would then require a number of replications of (fixed p_0) A -games that is about $200 \times p_A^A(0.5)/p_A^A(p_0) \approx 3 \times 10^{32}$. Assuming that 10^6 replications can be performed in a second by a super computer (which is overly optimistic), this estimation of $v_A^A(p_0)$ would still require not less than 9.5×10^{18} years! This means that it is indeed impossible to estimate in a reliable way the conditional variance curve for p close to p_0 so that Monte Carlo experiments *cannot* reveal the existence of the local maximum in p_0 . Similarly, without our theoretical analysis, there is no hope to learn about the degeneracy (respectively, nondegeneracy) of the limiting distribution of D conditional on a victory of A as $p \rightarrow 0$ (respectively, conditional on a victory of B as $p \rightarrow 1$).

We will not comment in detail the Monte Carlo results associated with the other figures. We just report that they again confirm our theoretical findings,

whenever possible, that is, whenever they are not associated with conditional results above events with small probabilities.

6. Conclusion and final comments

This paper provides a *complete* rally-level probabilistic description for games played under the side-out scoring system. It complements the previous main contributions from [7, 10] and [11] by adding to the well-known game-winning probabilities an exhaustive knowledge of the random duration of the game. This brings a much better understanding of the underlying process as a whole, as is demonstrated in Sections 2 to 4.

In this final section, we will mainly focus on the practical implications of our findings. For this, we may restrict to $(p_a, p_b) \in [0.4, 0.6] \times [0.4, 0.6]$, say, because players tend to be grouped according to strength. For such values of the rally-winning probabilities, our results show that the recent transition—in men's singles' Badminton—from the $n = 15$ side-out scoring system to the $n = 21$ rally-point one strongly affected the properties of the game. They indeed indicate that games played under the rally-point scoring system are much shorter than those played according to the side-out one, and that the uncertainty in the duration of the match is significantly reduced. Our results allow quantifying both effects. On the other hand, they show that game-winning probabilities are essentially the same for both scoring systems. It is then tempting to conclude (as in [11, 12]) that the outcomes of the games are barely influenced by the scoring system adopted. While this is strictly valid in the model, it is highly disputable under possible violations of the model. For instance, i.i.d.-ness (see page 2) may fail to hold for long games involving players with different fitness levels, a violation of the model under which scoring systems, through their impact on the duration of the games (see above), may significantly influence the outcomes of the games.

In practice, the results of this paper can be useful to many actors of the sport community. For the international sport federations playing with the idea of replacing the side-out scoring system with the rally-point one, our results could be used to tune n (i.e., the number of points to be scored to win a rally-point game) according to their wishes. For the sake of illustration, consider again the transition performed by the International Badminton Federation. Presumably, its objective was (i) to make the duration of the game more predictable and (ii) to ensure that the outcome of the matches would change as little as possible. If this was indeed the objective, then our results show that it has only been partially achieved: it is indeed easy to see that other choices of n would have been even better in that respect, the choice of $n = 27$ (see Figure 6(b) and (d)), being optimal. Moreover, this last choice would have affected the duration of the game much less than $n = 21$ (see Figure 6(c)), and thus would have

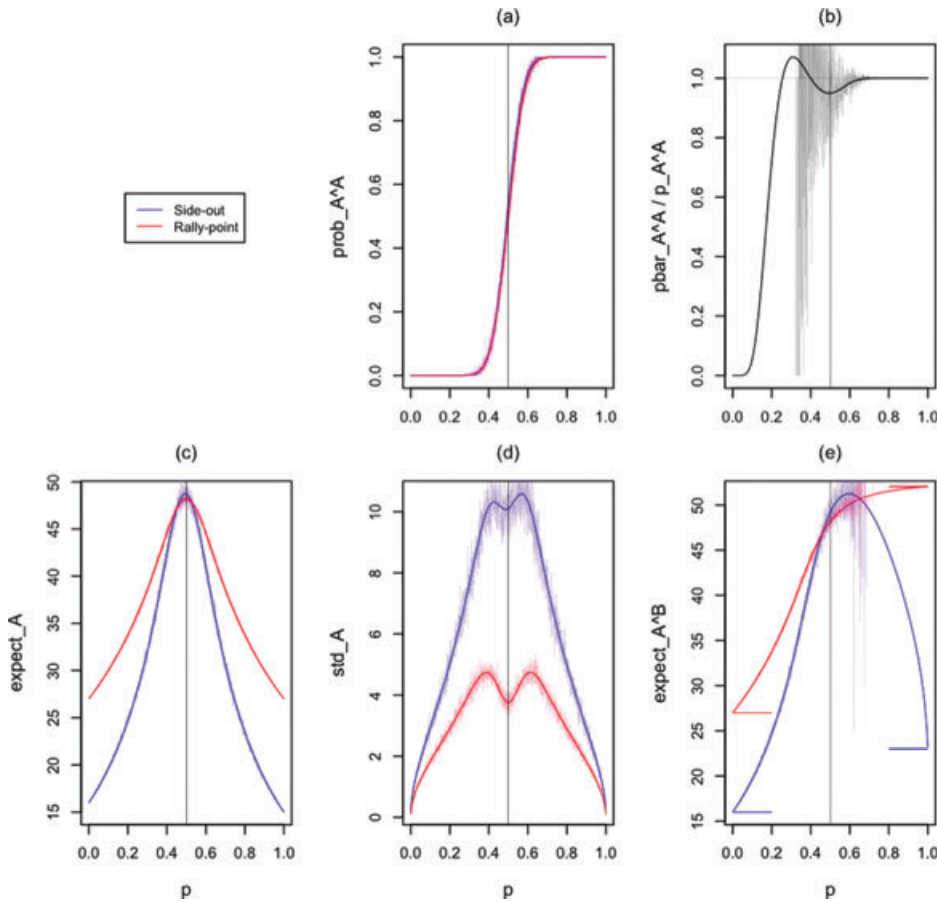


Figure 6. Subfigures (a)–(e) here report Subfigures (a)–(c), (f), and (e) from Figure 5 with the only difference that the rally-point scoring here is based on $n = 27$ (the side-out scoring is still based on $n = 15$).

made the outcome of the matches more robust to possible violations of the model.

For organizers of local tournaments played under the side-out scoring system, our results can be used to control, for any fixed number of planned matches, the time required to complete their events. Such a control over this random time, at any fixed tolerance level, can indeed be achieved in a quite direct way from our results on the duration of a game played under the side-out scoring system. Organizers can then deduce, at the corresponding tolerance level, the number of matches—hence the number of players—their events will be able to host. This of course concerns the sports that are still using this scoring system, such as racquetball and squash (for the latter, only in countries currently using the so-called English scoring system).

Finally, because our results provide a complete description of the duration process in the side-out scoring system, they also open the way to more efficient estimation of the rally-winning probabilities (p_a, p_b) there. However, a full discussion of this is beyond the scope of this paper, and is actually the topic of current research [8].

Appendix: proofs

A.1. Proofs of Lemma 1 and Theorem 1

In the Appendix, we simply write *interruptions* for *A-interruptions*.

Proof of Lemma 1. Clearly, $p_A^{\alpha,\beta,A}(r, j) = K_{r,j} p_a^\alpha p_b^\beta (q_a q_b)^{r+j}$, where $K_{r,j}$ is the number of ways of setting r interruptions and j exchanges in the sequence of rallies achieving the event under consideration. Regarding interruptions, we argue as in [2], and say those r interruptions should be put into the α possible spots (remember the last point should be won by A), while the β points scored by B should be distributed among those r interruptions—with at least one point scored by B in each interruption (so that there may be at most $r = \min(\alpha, \beta)$ interruptions). There are exactly $\binom{\alpha}{r} \binom{\beta-1}{r-1}$ ways to achieve this. As for the j exchanges, they may occur at any time and thus there are as many ways of placing j interruptions as there are distributions of j indistinguishable balls into $\alpha + \beta$ urns, i.e., $\binom{\alpha+\beta-1}{j}$. Summing up, we have proved that

$$p_A^{\alpha,\beta,A}(r, j) = \binom{\alpha + \beta - 1}{j} \binom{\alpha}{r} \binom{\beta - 1}{r - 1} p_a^\alpha p_b^\beta (q_a q_b)^{r+j},$$

with $r = \min(\beta, 1), \dots, \min(\alpha, \beta)$, $j \in \mathbb{N}$.

As for $p_A^{\alpha,\beta,B}(r, j)$, this probability is clearly of the form $L_{r,j} p_a^\alpha p_b^\beta q_a \times (q_a q_b)^{r+j-1}$. In this case, there are $\alpha + 1$ possible spots for the r interruptions. However, because B scores the last point, the sequence of rallies should end with an interruption. There are therefore $\binom{\alpha}{r-1}$ ways to insert the interruptions. Each interruption contains at least one point for B , so that $r \leq \min(\alpha + 1, \beta)$. The result follows by noting that there are $\binom{\beta-1}{r-1}$ ways of distributing the β points scored by B into those r interruptions, and by dealing with exchanges as for $p_A^{\alpha,\beta,A}(r, j)$. \square

Proof of Theorem 1. The result directly follows from Lemma 1 by writing $p_A^{\alpha,\beta,A} = \sum_{r,j} p_A^{\alpha,\beta,A}(r, j)$ and $p_A^{\alpha,\beta,B} = \sum_{r,j} p_A^{\alpha,\beta,B}(r, j)$ (where the sums are over all possible values of r and j in each case), and by using the equality $\sum_{j=0}^{\infty} \binom{m+j-1}{j} z^j = (1-z)^{-m}$ for any $z \in [0, 1)$. \square

A.2. Proofs of Theorems 2 and 3 and of Corollary 1

Proof of Theorem 2. First note that if A scores the last point in an A -game in which the score is α to β after j exchanges ($j \in \{0, 1, \dots\}$) and r interruptions ($r \in \{\gamma_0, \dots, \gamma_1\}$), then there have been $\alpha + \beta + 2(r + j)$ rallies. Conditioning on the number of interruptions and exchanges therefore yields

$$M_A^{\alpha, \beta, A}(t) = (p_A^{\alpha, \beta, A})^{-1} \sum_j \sum_r e^{t(\alpha + \beta + 2(r + j))} p_A^{\alpha, \beta, A}(r, j)$$

(where the sums are over all possible values of r and j in each case) and thus, from Lemma 1 and Theorem 1,

$$\begin{aligned} M_A^{\alpha, \beta, A}(t) &= \frac{e^{t(\alpha + \beta)} \sum_j (e^{2t} q)^j \binom{\alpha + \beta + j - 1}{j} \sum_r e^{2tr} \binom{\alpha}{r} \binom{\beta - 1}{r - 1} q^r}{(1 - q)^{-(\alpha + \beta)} \sum_r \binom{\alpha}{r} \binom{\beta - 1}{r - 1} q^r} \\ &= ((1 - q)e^t)^{\alpha + \beta} \left(\sum_j (e^{2t} q)^j \binom{\alpha + \beta + j - 1}{j} \right) \\ &\quad \times \left(\sum_r e^{2tr} W_A^{\alpha, \beta, A}(q, r) \right). \end{aligned}$$

The first claim of Theorem 2 follows.

For the second claim, it suffices to note that if B scores the last point in an A -game in which the score is of α to β after j exchanges ($j \in \{0, 1, \dots\}$) and r interruptions ($r \in \{1, \dots, \gamma_2 + 1\}$), then the number of rallies equals $\alpha + \beta + 2(r - 1 + j) + 1$; the computations above then hold with only minor changes. \square

Proof of Corollary 1. Taking first and second derivatives of the moment generating functions yields the expectations and variances given in Corollary 1. Moreover it can easily be seen that derivatives of the expected values with respect to q are positive by using the Cauchy–Schwarz inequality, and thus the latter are strictly monotone increasing in q . \square

Proof of Theorem 3. The change of variables $z = e^t$ in the moment generating functions given in Theorem 2 immediately yields the probability generating functions. If $\beta = 0$, the latter is already in the form of an infinite series $G_A^{\alpha, 0, A}(z) = \sum_{j=0}^{\infty} (1 - q)^\alpha q^j \binom{\alpha + j - 1}{j} z^{\alpha + 2j}$. If $\beta > 0$, we have

$$G_A^{\alpha, \beta, A}(z) = (1 - q)^{\alpha + \beta} z^{\alpha + \beta} \sum_{j=0}^{\infty} K_j z^{2j} \sum_{r=1}^{\gamma_1} W_r z^{2r},$$

where $K_j = q^{j(\alpha+\beta+j-1)}$ and $W_r = W_A^{\alpha,\beta,A}(q, r)$. This double sum satisfies

$$\sum_{j=0}^{\infty} K_j \sum_{r=1}^{\gamma_1} W_r z^{2(j+r)} = \sum_{j=1}^{\gamma_1} z^{2j} \left(\sum_{l=0}^{j-1} K_l W_{j-l} \right) + \sum_{j=\gamma_1+1}^{\infty} z^{2j} \left(\sum_{l=j-\gamma_1}^{j-1} K_l W_{j-l} \right).$$

The same arguments are readily adapted to $G_A^{\alpha,\beta,B}(z)$, and Theorem 3 follows. \square

A.3. The distribution of the number of rallies, in the no-server model, for extreme rally-winning probabilities

As announced in Section 4.2, we determine here the limiting behavior of the number of rallies D , in the no-server model, for $p \rightarrow 0$ and $p \rightarrow 1$, conditional on the winner of the A -game considered. We start with the limit under almost sure events, that is, limits as $p \rightarrow 1$ (respectively, $p \rightarrow 0$) for the distribution of D conditional on a victory of A (respectively, of B).

PROPOSITION 1. *Let, for the side-out scoring system, $t \mapsto M_A^C(t) = \mathbb{E}[e^{tD} \mid E^C, S = A]$, $C \in \{A, B\}$, be the moment generating function of D conditional on the event $E^C \cap [S = A]$. Denote by $t \mapsto \bar{M}_A^C(t) = \mathbb{E}[e^{tD} \mid \bar{E}^C, S = A]$, $C \in \{A, B\}$, the corresponding moment generating function for the rally-point system. Then, (i) as $p \rightarrow 1$, $M_A^A(t) \rightarrow e^{nt}$ and $\bar{M}_A^A(t) \rightarrow e^{nt}$; (ii) as $p \rightarrow 0$, $M_A^B(t) \rightarrow e^{(n+1)t}$ and $\bar{M}_A^B(t) \rightarrow e^{nt}$.*

Proof. (i) By conditioning, we get $M_A^A(t) = \sum_{k=0}^{n-1} M_A^{n,k,A}(t) p_A^{n,k,A} / p_A^A$. It is easy to check that $\lim_{p \rightarrow 1} p_A^{n,k,A} / p_A^A = \delta_{k,0}$ and that $\lim_{p \rightarrow 1} M_A^{n,k,A}(t) = e^{(n+k)t}$. Hence $\lim_{p \rightarrow 1} M_A^A(t) = e^{nt}$. Likewise, $\bar{M}_A^A(t) = \sum_{k=0}^{n-1} e^{(n+k)t} \bar{p}_A^{n,k,A} / \bar{p}_A^A$. Again, it is easy to check that $\lim_{p \rightarrow 1} \bar{p}_A^{n,k,A} / \bar{p}_A^A = \delta_{k,0}$. Hence, we indeed have $\bar{M}_A^A(t) \rightarrow e^{nt}$. (ii) The proof is similar, and thus left to the reader. \square

COROLLARY 2. *(i) As $p \rightarrow 1$, $(e_A^A, v_A^A) \rightarrow (n, 0)$ and $(\bar{e}_A^A, \bar{v}_A^A) \rightarrow (n, 0)$, so that, conditional on a victory of A in an A -game, $D \xrightarrow{p} n$, irrespective of the scoring system; (ii) as $p \rightarrow 0$, $(e_A^B, v_A^B) \rightarrow (n+1, 0)$ and $(\bar{e}_A^B, \bar{v}_A^B) \rightarrow (n, 0)$, so that, conditional on a victory of B in an A -game, $D \xrightarrow{p} n+1$ (respectively, n) for the side-out (respectively, rally-point) scoring system.*

As shown by Proposition 1 and Corollary 2, the situation is here very clear. In each of the four cases considered, only one trajectory is possible, namely that for which all rallies in the game will be won by the winner of the game.

Next we derive the limiting conditional distribution of D under events which occur with zero probability, that is, limits as $p \rightarrow 1$ (respectively, $p \rightarrow 0$) for

the distribution of D conditional on a victory of B (respectively, of A). Our conclusions are much more surprising.

PROPOSITION 2. *Let $m(t) := \sum_{k=0}^{n-1} e^{(n+k)t} \binom{n+k-1}{k} / \sum_{k=0}^{n-1} \binom{n+k-1}{k}$. Then, (i) as $p \rightarrow 0$, $M_A^A(t) \rightarrow e^{nt}$ and $\bar{M}_A^A(t) \rightarrow m(t)$; (ii) as $p \rightarrow 1$, $M_A^B(t) \rightarrow (e^{(n+1)t} - e^{(2n+1)t}) / (n(1 - e^t))$ and $\bar{M}_A^B(t) \rightarrow m(t)$. In particular, as $p \rightarrow 1$, the limiting distribution of D conditional on the event $E^B \cap [S = A]$ is uniform over the set $\{n + 1, \dots, 2n\}$.*

Proof. We first prove the assertions for the rally-point scoring system. In this case, $\bar{M}_A^A(t) = \sum_{k=0}^{n-1} e^{(n+k)t} \bar{p}_A^{n,k,A} / \bar{p}_A^A$. Now, from Remark 1 it is immediate that $\lim_{p \rightarrow 0} p_A^{n,k,A} / p_A^A = \binom{n+k-1}{k} / \sum_{k=0}^{n-1} \binom{n+k-1}{k}$, which proves the claim for $\bar{M}_A^A(t)$ (hence, by symmetry, also for $\bar{M}_A^B(t)$).

Next consider the assertions for the side-out scoring system. First note that, as before, $M_A^A(t) = \sum_{k=0}^{n-1} M_A^{n,k,A}(t) p_A^{n,k,A} / p_A^A$ and $M_A^B(t) = \sum_{k=0}^{n-1} M_A^{k,n,B}(t) p_A^{k,n,B} / p_A^B$. Now fix $k \in \{0, \dots, n-1\}$. Using Theorem 1, one readily shows that

$$\lim_{p \rightarrow 0} p_A^{n,k,A} / p_A^A = \delta_{k,0} \quad \text{and} \quad \lim_{p \rightarrow 1} p_A^{k,n,B} / p_A^B = 1/n.$$

Combining these results and the definitions of the moment generating functions, it is then straightforward to show that

$$\lim_{p \rightarrow 0} M_A^{n,k,A}(t) = e^{(n+k)t} \quad \text{and} \quad \lim_{p \rightarrow 1} M_A^{k,n,B}(t) = e^{(n+k+1)t}.$$

The claim follows. \square

COROLLARY 3. (i) As $p \rightarrow 0$, $(e_A^A, v_A^A) \rightarrow (n, 0)$ and $(\bar{e}_A^A, \bar{v}_A^A) \rightarrow (\frac{2n^2}{n+1}, \frac{2n^2(n-1)}{(n+1)^2(n+2)})$; as $p \rightarrow 1$, $(e_A^B, v_A^B) \rightarrow (\frac{3n+1}{2}, \frac{(n-1)^2}{12})$ and $(\bar{e}_A^B, \bar{v}_A^B) \rightarrow (\frac{2n^2}{n+1}, \frac{2n^2(n-1)}{(n+1)^2(n+2)})$.

It is remarkable that we can again give a complete description of the “distribution of the process” (by this, we mean that we can again list all trajectories of rallies leading to the event considered, and give, for each such trajectory, its probability). Consider first the side-out scoring system. For victories of A , the situation is very clear: Corollary 2 indeed yields that, conditional on a victory of A in an A -game, $D \xrightarrow{P} n$ as $p \rightarrow 0$, which implies that the only possible trajectory of rallies is the one for which all rallies in the game are won by A . Turn then to victories of B . There, we obtained in the proof of Proposition 2 that all scores (k, n) are equally likely. It is actually easy to show that, conditional on $E^{k,n,B} \cap [S = A]$, $D \xrightarrow{P} n + k + 1$ as $p \rightarrow 1$. This implies that there are exactly n equally likely trajectories: A first scores

k points, then loses his/her serve, before B scores n (miraculous) points and wins the game ($k = 0, \dots, n - 1$).

Consider finally the rally-point system. In this case, it is sufficient to study the distribution of the scores after victories of A (when $p \rightarrow 0$) because the number of rallies is a function of the scores only, and because the conclusions will, by symmetry, be identical for victories of B (when $p \rightarrow 1$). Clearly, for any fixed $k \in \{0, 1, \dots, n - 1\}$, there are exactly $\binom{n+k-1}{k}$ trajectories leading to the score (n, k) , and those trajectories are equally likely. Each such trajectory will then occur with probability $1 / \sum_{k=0}^{n-1} \binom{n+k-1}{k}$, because, as we have seen in the proof of Proposition 2, the score (n, k) occurs with probability $\binom{n+k-1}{k} / \sum_{k=0}^{n-1} \binom{n+k-1}{k}$. These considerations provide the whole distribution of the process: there are $\sum_{k=0}^{n-1} \binom{n+k-1}{k}$ equally likely possible trajectories, namely the ones we have just considered. The exact limiting distribution of D can of course trivially be computed from this.

Acknowledgments

The research of Davy Paindaveine (who is also member of ECORE, the association of CORE and ECARES) is supported by an A.R.C. contract of the Communauté Française de Belgique; and that of Yvik Swan is supported by a Mandat de Chargé de Recherche from the Fonds National de la Recherche Scientifique, Communauté française de Belgique. The authors thank an anonymous referee for insightful comments and suggestions that led to a significant improvement of the paper.

References

1. C. L. ANDERSON, Note on the advantage of first serve, *J. Combin. Theory Ser. A* 23:363 (1977).
2. F. J. G. M. KLAASSEN and J. R. MAGNUS, Forecasting the winner of a tennis match, *European J. Oper. Res.* 148:257–267 (2003).
3. P. K. NEWTON and J. B. KELLER, Probability of winning at tennis I. Theory and data, *Stud. Appl. Math.* 114:241–269 (2005).
4. R. S. SCHULMAN and M. A. HAMDAN, A probabilistic model for table tennis, *Canad. J. Statist.* 5:179–186 (1977).
5. P. K. NEWTON and K. ASLAM, Monte Carlo Tennis, *SIAM Rev.* 48:722–742 (2006).
6. B. P. HSI and D. M. BURICH, Games of two players, *J. Roy. Statist. Soc. Ser. C* 20:86–92 (1971).
7. M. J. PHILLIPS, Sums of random variables having the modified geometric distribution with application to two-person games, *Adv. Appl. Probab.* 10:647–665 (1978).
8. J. B. KELLER, Probability of a shutout in racquetball, *SIAM Rev.* 26:267–268 (1984).

9. D. J. MARCUS, Probability of winning a game of racquetball, *SIAM Rev.* 27:443–444 (1985).
10. D. STRAUSS and B. C. ARNOLD, The rating of players in racquetball tournaments, *J. Roy. Statist. Soc. Ser. C* 36:163–173 (1987).
11. J. SIMMONS, A probabilistic model of squash: strategies and applications, *J. Roy. Statist. Soc. Ser. C* 38:95–110 (1989).
12. D. F. PERCY, A mathematical analysis of badminton scoring systems, *J. Oper. Res. Soc.* 60:63–71 (2009).
13. J. RENICK, Optional strategy at decision points, *Res. Quart.* 47:562–568 (1976).
14. J. RENICK, Tie point strategy in badminton and international squash, *Res. Quart.* 48:492–498 (1977).

UNIVERSITÉ LIBRE DE BRUXELLES, BELGIUM

(Received August 1, 2010)