On the length of winning a tennis game using random walks

TOSHIO NAKATA

Abstract

This article is a follow-up to a previous *Gazette* article that examined the length of a winning tennis game, as studied by Cooper and Kennedy (2021). Let X denote a random variable representing the length of a tennis game under standard rules. Using random walks, we present simple expressions for the probability distribution $\mathrm{P}(X = k)$, the expectation $\mathrm{E}(X)$, and the variance $\mathrm{Var}(X)$. In addition, we derive the expected values of the maximum and minimum lengths across n games, together with asymptotic approximations for these quantities.

Keywords: tennis game, random walk, Markov chain, absorption time

1. *Introduction*

Consider players A and B playing tennis. For $0 \leq p \leq 1$, let p denote the probability that player A wins the next point. A game ends when either player A or B scores at least four points with a two-point margin. The rules for points and games are assumed to follow the standard scoring system, as described in [1, Section 2] and Liu [2, Section 1]. Stewart [3, Section 2.4, p. 23] gave

$$\mathrm{P}(\text{player A wins a game}) = \frac{15p^4 - 34p^5 + 28p^6 - 8p^7}{1 - 2p + 2p^2}.$$

Liu [2, Section 2.5, p. 159] also showed it using random walks. Let X be a random variable representing the length of the game. Cooper and Kennedy [1, Section 3, p. 494] proved

$$\mathrm{E}(X) = \frac{4(1 - p + p^2 + 6p^3 - 18p^4 + 18p^5 - 6p^6)}{1 - 2p + 2p^2}. \tag{1}$$

However, the calculations seem complicated because the Cayley-Hamilton theorem for large matrices is used. In this article, following [2], we present a simpler approach based on random walks and provide simple expressions for the distribution $\mathrm{P}(X = k)$, the expectation $\mathrm{E}(X)$, and the variance $\mathrm{Var}(X)$. Letting $q = 1 - p$ and

$$r = pq, \tag{2}$$

we show that not only (1) but also $\mathrm{Var}(X)$ can be expressed in terms of r rather than p (see Theorem 1). Moreover, we provide the expectations of the maximum and minimum lengths for n games (see Theorem 2). Finally, we explore simple asymptotic estimates for these expectations (see Theorem 3).

2. *The distribution of the length of winning a game*

To begin, we examine the properties of the parameters p, q, and r. Since $0 \le p \le 1, p + q = 1$, and (2), it follows that

$$\begin{cases} p^2 + q^2 = 1 - 2r, \\ p^3 + q^3 = 1 - 3r, \\ p^4 + q^4 = 1 - 4r + 2r^2, \end{cases} \tag{3}$$

and

$$0 \le r \le \frac{1}{4}. \tag{4}$$

While the points in tennis are traditionally expressed as $\{0, 15, 30, 40\}$, we simplify them to $\{0, 1, 2, 3\}$ for clarity. We consider a pair of points for players A and B as a state (a, b) for $a, b \in \{0, 1, 2, 3\}$, and we introduce additional states "A's game", "Ad-in", "Deuce", "Ad-out", and "B's game", which appear as ovals in Figure 1. Each state, except "A's game" and "B's game", transitions with probability p or q as shown in Figure 1. We investigate a Markov chain with two absorbing barriers "A's game" and "B's game". In this setting, X can be regarded as the absorption time with the initial state $(0, 0)$. For simplicity, denoting the states that are ovals in Figure 1 by

$$(\text{A's Game}, \text{Ad-in}, \text{Deuce}, \text{Ad-out}, \text{B's Game}) = (4, 3, 2, 1, 0),$$

we treat it as the *gambler's ruin problem* (cf. Grimmett and Stirzaker [4, Section 1.7.4, p. 18]). The following lemma is used to calculate the probability of winning a game of tennis.

Lemma 1: Let $\eta_1, \eta_2, \dots$ be independent and identically distributed random variables with $\mathrm{P}(\eta_1 = 1) = p$ and $\mathrm{P}(\eta_1 = -1) = q$. We consider a random walk $Y_n = \sum_{k=1}^{n} \eta_k$. Let a_n, b_n, and c_n be the probabilities that the random

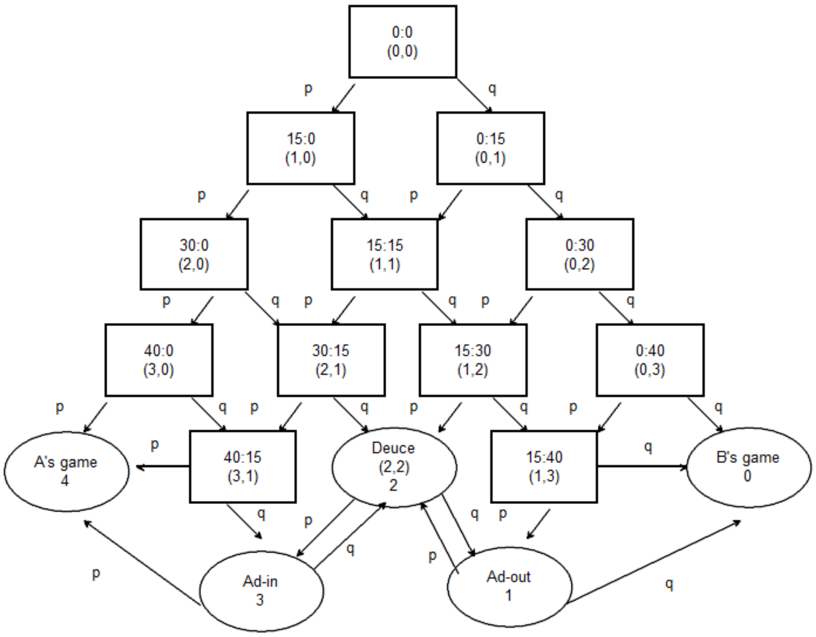

Figure 1: The transition diagram for the tennis game

walk first visits $\{0, 4\}$ at the nth step when the initial states are $1, 2$, and 3, respectively. More precisely,

$$\begin{cases} a_n = \mathrm{P}(Y_n \in \{0,4\}, Y_i \notin \{0,4\}, 0 \le i < n | Y_0 = 1), \\ b_n = \mathrm{P}(Y_n \in \{0,4\}, Y_i \notin \{0,4\}, 0 \le i < n | Y_0 = 2), \\ c_n = \mathrm{P}(Y_n \in \{0,4\}, Y_i \notin \{0,4\}, 0 \le i < n | Y_0 = 3). \end{cases}$$

Then it follows that

$$b_{2k} = (1-2r)(2r)^{k-1} \quad \text{for } k = 1, 2, \dots. \tag{5}$$

Proof: When $k = 1$, (5) holds because of (3). Now, assume that $k \ge 2$. By the first step analysis, we obtain

$$\begin{cases} a_{2k-1} & = \begin{cases} pb_{2k-2} & \text{for } k \ge 2, \\ q & \text{for } k = 1, \end{cases} \\ b_{2k} & = \begin{cases} pc_{2k-1} + qa_{2k-1} & \text{for } k \ge 2, \\ p^2 + q^2 & \text{for } k = 1, \end{cases} \\ c_{2k-1} & = \begin{cases} qb_{2k-2} & \text{for } k \ge 2, \\ p & \text{for } k = 1. \end{cases} \end{cases}$$

Therefore, it follows that

$$b_{2k} = 2pqb_{2(k-1)} = (2pq)^2 b_{2(k-2)} = \cdots = (2pq)^{k-1} b_2 = (2pq)^{k-1}(p^2+q^2).$$

Applying (2) and (3) to b_{2k} proves (5).

Remark 1: We note that (5) can be interpreted as a geometric distribution with parameter $2r$ when states 1 and 3 are identified, as well as states 0 and 4.

Using Figure 1 and Lemma 1, we see the distribution of X as follows.

Lemma 2: An arbitrary tennis game ends with an even number of steps, except when $X = 5$. More precisely, we have $\mathrm{P}(X \in \{2m : m = 2, 3, 4, \ldots\} \cup \{5\}) = 1$ and

$$\mathrm{P}(X = i) = \begin{cases} 1 - 4r + 2r^2 & \text{for } i = 4, \\ 4r(1-3r) & \text{for } i = 5, \\ 5r(1-2r)(2r)^{m-2} & \text{for } i = 2m,\ m = 3, 4, 5, \ldots. \end{cases} \tag{6}$$

Proof: Figure 1 yields $X \geq 4$. Let ξ_{A4} be a point of player A at the 4th step. We then have $\mathrm{P}(\xi_{A4} = i) = \binom{4}{i} p^i q^{4-i}$ for $i \in \{0, \ldots, 4\}$, and the state of the Markov chain is $(\xi_{A4}, 4 - \xi_{A4})$. By using ξ_{A4}, all games are classified as one of the following.

- A case where a game ends at the 4th step: Since $\{X = 4\}$ is the event of four consecutive points, we have

  $$\mathrm{P}(X = 4) = \mathrm{P}(\xi_{A4} \in \{0, 4\}) = p^4 + q^4 \overset{(3)}{=} 1 - 4r + 2r^2.$$

- A case where a game does *not* end at the 4th step: The Markov chain is in $(1, 3), (3, 1)$, or $(2, 2)$ at the 4th step.
  - A case where a game ends at the 5th step: Since it is in $(1, 3)$ or $(3, 1)$, we have

    $$\begin{aligned} \mathrm{P}(X = 5) &= \mathrm{P}(\xi_{A4} = 1)\mathrm{P}(X = 5 | \xi_{A4} = 1) \\ &\quad + \mathrm{P}(\xi_{A4} = 3)\mathrm{P}(X = 5 | \xi_{A4} = 3) \\ &= 4pq(p^3 + q^3) \overset{(3)}{=} 4r(1 - 3r). \end{aligned}$$

- A case where a game does *not* end at the 5th step: The game ends with an even number of steps for the following reasons.
  - If it is in "Deuce $(2,2)$" at the 4th step, then (5) whose period is two is applicable.
  - If it is in $(1,3)$ or $(3,1)$ at the 4th step, then either the game ends at the 6th step or it visits "Deuce $(3,3)$" at the 6th step.

  When $m \geq 4$, since

$$\begin{cases} \mathrm{P}(X = 2m | \xi_{A4} = 1) = p^2 b_{2(m-3)} \\ \mathrm{P}(X = 2m | \xi_{A4} = 2) = b_{2(m-2)} \\ \mathrm{P}(X = 2m | \xi_{A4} = 3) = q^2 b_{2(m-3)}, \end{cases}$$

  we have

$$\begin{aligned} \mathrm{P}(X = 2m) &= \sum_{i=1}^{3} \mathrm{P}(\xi_{A4} = i)\mathrm{P}(X = 2m | \xi_{A4} = i) \\ &= 4pq^3 \times p^2 b_{2(m-3)} + 6p^2q^2 \times b_{2(m-2)} + 4p^3q \times q^2 b_{2(m-3)} \\ &= 5pq(p^2 + q^2)(2pq)^{m-2} \overset{(3)}{=} 5r(1-2r)(2r)^{m-2}, \end{aligned}$$

  which proves (6). The case of $m = 3$ of (6) can be proved by direct calculation.

Lemma 2 provides both $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

Theorem 1: The expectation and the variance of the length of the tennis game X are

$$\begin{aligned} \mathrm{E}(X) &= \frac{4(1 - r + 6r^3)}{1 - 2r}, \\ \mathrm{Var}(X) &= \frac{4r(1 - r + 36r^2 - 60r^3 - 144r^5)}{(1-2r)^2}, \end{aligned}$$

respectively. Moreover, we have the following.

- When $r = 0$, namely, $p \in \{0, 1\}$ holds, $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ attain their minimum values 4 and 0, respectively.
- When $r = \frac{1}{4}$, namely, $p = q = \frac{1}{2}$ holds, $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ attain their maximum values $\frac{27}{4} = 6.75$ and $\frac{123}{16} = 7.6875$, respectively.

Proof: By noting $0 \le 2r \le \frac{1}{2}$, a direct calculation yields

$$\begin{aligned}
\mathrm{E}(X) &= 4(1-4r+2r^2)+20r(1-3r)+5r(1-2r)\sum_{n=3}^{\infty} 2n(2r)^{n-2} \\
&= 4+4r-52r^2+5(1-2r)\left\{\sum_{n=1}^{\infty} n(2r)^{n-1}-1-4r\right\} \\
&= \frac{4(1-r+6r^3)}{1-2r}.
\end{aligned}$$

Since

$$\begin{aligned}
& \mathrm{E}(X(X-2)) \\
={}& 8(1-4r+2r^2)+60r(1-3r)+5r(1-2r)\sum_{n=3}^{\infty} 2n(2n-2)(2r)^{n-2} \\
={}& 8+28r-164r^2+20r(1-2r)\left\{\sum_{n=2}^{\infty} n(n-1)(2r)^{n-2}-2\right\} \\
={}& \frac{4(2-r-r^2+72r^3-84r^4)}{(1-2r)^2},
\end{aligned}$$

the variance is

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2)-\{\mathrm{E}(X)\}^2 = \mathrm{E}(X(X-2))-\mathrm{E}(X)(\mathrm{E}(X)-2) \\
&= \frac{4r(1-r+36r^2-60r^3-144r^5)}{(1-2r)^2}.
\end{aligned}$$

Putting $f(r)=\mathrm{E}(X)$, we see $f'(r)=\frac{4\left\{1+6r^2(3-4r)\right\}}{(1-2r)^2}>0$ for $0 \le r \le \frac{1}{4}$. Hence the minimum is $f(0)=0$ and the maximum is $f\left(\frac{1}{4}\right)=\frac{27}{4}$. Similarly, putting $g(r)=\mathrm{Var}(X)$ we obtain

$$g'(r)=\frac{4\left(1152r^6-864r^5+240r^4-312r^3+108r^2+1\right)}{(1-2r)^3}.$$

Since the polynomial of r in the brackets of the numerator is

$$1152r^6-864r^5+240r^4-312r^3+108r^2+1 \ge 1+108^2(1-4r)+240r^4(1-4r)+1152r^6>0,$$

we see $g'(r)>0$ for $0 \le r \le \frac{1}{4}$. Hence the minimum is $g(0)=0$ and the maximum is $g\left(\frac{1}{4}\right)=\frac{123}{16}$.

Remark 2: If we use p instead of r in $\mathrm{E}(X)$ in Theorem 1, it becomes equivalent to (1). The fact that the maximum of $\mathrm{E}(X)$ is 6.75 has been shown in [1, p. 494].

3. *The maximum and minimum lengths for n games*

We examine the maximum and minimum lengths of the games when the game is repeated independently n times. Let $X_{1,n}$ and $X_{n,n}$ denote the minimum and maximum values, respectively, among the lengths of n independent games. Namely, letting $X_1, X_2, \ldots, X_n$ be the independent copies of the random variable X, they are defined as follows:

$$X_{1,n} = \min\{X_1, X_2, \ldots, X_n\} \quad \text{and} \quad X_{n,n} = \max\{X_1, X_2, \ldots, X_n\}.$$

To see the distributions of $X_{1,n}$ and $X_{n,n}$, we calculate the tail probability of X. It follows from (6) that

$$\mathrm{P}(X \geq i) = \begin{cases} 1 & \text{for } i \in \{0, \ldots, 4\}, \\ 4r - 2r^2 & \text{for } i = 5, \\ 10r^2 & \text{for } i = 6, \\ 5r(2r)^{m-2} & \text{for } i \in \{2m-1, 2m\},\ m = 4, 5, 6, \ldots. \end{cases} \tag{7}$$

The distribution is light-tailed and resembles a geometric distribution.

In general, if ξ is a nonnegative integer valued summable random variable, then we get

$$\mathrm{E}(\xi) = \sum_{k=0}^{\infty} \mathrm{P}(\xi > k) = \sum_{k=1}^{\infty} \mathrm{P}(\xi \geq k) \tag{8}$$

(see [4, Problem 3.11.13 (a), p. 90]). By applying (7) to (8), we can derive $\mathrm{E}(X)$ more easily in the following way.

$$\begin{aligned} \mathrm{E}(X) &= \sum_{k=1}^{\infty} \mathrm{P}(X \geq k) = 4 + (4r - 2r^2) + 10r^2 + 2 \times 5r \sum_{m=4}^{\infty} (2r)^{m-2} \\ &= 4 + 4r + 8r^2 + \frac{10r(2r)^2}{1 - 2r} = \frac{4(1 - r + 6r^3)}{1 - 2r}. \end{aligned}$$

We derive the probability distributions and expectations of $X_{1,n}$ and $X_{n,n}$.

Theorem 2: The distributions of the minimum length $X_{1,n}$ and the maximum length $X_{n,n}$ for n games are

$$\begin{aligned}
&\mathrm{P}(X_{1,n} = i)\\
= &\left\{\begin{array}{ll} 1-(4r-2r^2)^n & \text{for } i=4,\\ (4r-2r^2)^n-(10r^2)^n & \text{for } i=5,\\ \{5r(2r)^{m-2}\}^n-\{5r(2r)^{m-1}\}^n & \text{for } i=2m,\ m=3,4,5,\ldots, \end{array}\right. \qquad (9)\\
&\mathrm{P}(X_{n,n} = i)\\
= &\left\{\begin{array}{ll} (1-4r+2r^2)^n & \text{for } i=4,\\ (1-10r^2)^n-(1-4r+2r^2)^n & \text{for } i=5,\\ \{1-5r(2r)^{m-1}\}^n-\{1-5r(2r)^{m-2}\}^n & \text{for } i=2m,\ m=3,4,5,\ldots, \end{array}\right. \qquad (10)
\end{aligned}$$

respectively. Moreover, the expectations are

$$\begin{aligned}
\mathrm{E}(X_{1:n}) &= 4+(4r)^n\left(1-\frac{r}{2}\right)^n+(10r^2)^n\left(1+\frac{2(2r)^n}{1-(2r)^n}\right), \qquad (11)\\
\mathrm{E}(X_{n:n}) &= 6-(1-4r+2r^2)^n-(1-10r^2)^n\\
&\quad +2\sum_{m=0}^{\infty}\left[1-\left\{1-5r(2r)^{m+2}\right\}^n\right]. \qquad (12)
\end{aligned}$$

Proof: Let $i \geq 0$ be an integer. Since

$$\begin{aligned}
\mathrm{P}(X_{1,n} \geq i) &= \mathrm{P}(X_1 \geq i, \ldots, X_n \geq i) = \{\mathrm{P}(X \geq i)\}^n\\
&\overset{(7)}{=} \left\{\begin{array}{ll} 1 & \text{for } i \in \{0,\ldots,4\},\\ (4r-2r^2)^n & \text{for } i=5,\\ (10r^2)^n & \text{for } i=6,\\ \{5r(2r)^{m-2}\}^n & \text{for } i \in \{2m-1, 2m\},\ m=4,5,6,\ldots. \end{array}\right.
\end{aligned}$$

and $\mathrm{P}(X_{1,n} = i) = \mathrm{P}(X_{1,n} \geq i) - \mathrm{P}(X_{1,n} \geq i+1)$ for $i \geq 4$, we have (9). It follows that

$$\begin{aligned}
\mathrm{E}(X_{1:n}) &\overset{(8)}{=} \sum_{k=1}^{\infty}\mathrm{P}(X_{1:n} \geq k) = 4+(4r-2r^2)^n+(10r^2)^n+2\times(5r)^n\sum_{m=4}^{\infty}\{(2r)^n\}^{m-2}\\
&= 4+(4r-2r^2)^n+(10r^2)^n+\frac{2(20r^3)^n}{1-(2r)^n}\\
&= 4+(4r)^n\left(1-\frac{r}{2}\right)^n+(10r^2)^n\left(1+\frac{2(2r)^n}{1-(2r)^n}\right).
\end{aligned}$$

From (7) we have

$$\mathrm{P}(X \le i) = \begin{cases} 0 & \text{for } i \in \{0,1,2,3\}, \\ 1-4r+2r^2 & \text{for } i = 4, \\ 1-10r^2 & \text{for } i = 5, \\ 1-5r(2r)^{m-1} & \text{for } i \in \{2m, 2m+1\},\ m = 3,4,\ldots. \end{cases}$$

Since

$$\begin{aligned} \mathrm{P}(X_{n,n} \le i) &= \mathrm{P}(X_1 \le i, \ldots, X_n \le i) = \{\mathrm{P}(X \le i)\}^n \\ &= \begin{cases} 0 & \text{for } i \in \{0,1,2,3\}, \\ (1-4r+2r^2)^n & \text{for } i = 4, \\ (1-10r^2)^n & \text{for } i = 5, \\ \{1-5r(2r)^{m-1}\}^n & \text{for } i \in \{2m, 2m+1\},\ m = 3,4,\ldots. \end{cases} \end{aligned}$$

and $\mathrm{P}(X_{n,n} = i) = \mathrm{P}(X_{n,n} \le i) - \mathrm{P}(X_{n,n} \le i-1)$ for $i \ge 4$, we have (10). Similarly, it follows that

$$\begin{aligned} \mathrm{E}(X_{n:n}) &\overset{(8)}{=} \sum_{i=0}^{\infty} \{1 - \mathrm{P}(X_{n:n} \le i)\} \\ &= 6 - (1-4r+2r^2)^n - (1-10r^2)^n + 2\sum_{m=0}^{\infty} \left[1 - \left\{1 - 5r(2r)^{m+2}\right\}^n\right]. \end{aligned}$$

We provide simple asymptotic evaluations of both $\mathrm{E}(X_{1:n})$ and $\mathrm{E}(X_{n:n})$.

Theorem 3: Suppose $0 < p < 1$. Then we have

$$\begin{aligned} \mathrm{E}(X_{1:n}) &= 4 + o(1), && (13) \\ \mathrm{E}(X_{n:n}) &= \frac{2\log n}{-\log(2r)} + O(1), && (14) \end{aligned}$$

where $o(\cdot)$ and $O(\cdot)$ denote the standard *Landau symbols*, with $o(1)$ converges to 0 as $n \to \infty$, and $O(1)$ indicating boundedness.

Proof: Note that $0 < r \le \frac{1}{4}$ because of $0 < p < 1$ and (4). Let $n \ge 1$ be an integer. It follows that

$$0 < 4r \le 1, \quad \frac{7}{8} \le 1 - \frac{r}{2} < 1, \quad 0 < 10r^2 \le \frac{5}{8}, \quad \text{and} \quad 0 < \frac{2(2r)^n}{1-(2r)^n} \le 2.$$

Applying them to (11) yields (13). Next, we prove (14) from (12). Note that $6-(1-4r+2r^2)^n-(1-10r^2)^n$ is bounded. Hence, if

$$\begin{aligned} &\frac{\log n}{-\log(2r)} - 3 + \frac{\log(2.5)}{-\log(2r)} \leq \sum_{m=0}^{\infty} \left[1 - \left\{1 - 5r(2r)^{m+2}\right\}^n\right] \\ &\leq \frac{\log n}{-\log(2r)} - \frac{1 - 10r^2 - \gamma}{-\log(2r)} \end{aligned} \tag{15}$$

holds, then (14) is shown, where $\gamma = 0.5772\ldots$ is the *Euler-Mascheroni constant*. In the following, we check (15). Let $m \geq 0$ be an integer. Putting

$$\lambda = -\log(2r) \geq \log 2 \quad \text{and} \quad 0 < c = \log\frac{5}{2} < 1,$$

we have $2r = e^{-\lambda}$, $5r = e^{-\lambda+c}$, and $1-\left\{1-5r(2r)^{m+2}\right\}^n = 1-\left(1-e^{-\lambda(m+3)+c}\right)^n$. Since $x \mapsto 1-\left(1-e^{-\lambda x+c}\right)^n$ is strictly decreasing for $x \geq 0$, it turns out that

$$1-\left(1-e^{-\lambda x+c}\right)^n \leq 1-\left(1-e^{-\lambda(m+3)+c}\right)^n \leq 1-\left(1-e^{-\lambda(x-1)+c}\right)^n \quad \text{for } m+3 \leq x \leq m+4.$$

Integrating over x from $m+3$ to $m+4$, and summing for $m \geq 0$, yields

$$\begin{aligned} &\int_3^{\infty} 1-\left(1-e^{-\lambda x+c}\right)^n dx \leq \sum_{m=0}^{\infty} \left\{1-\left(1-e^{-\lambda(m+3)+c}\right)^n\right\} \\ &\leq \int_2^{\infty} 1-\left(1-e^{-\lambda x+c}\right)^n dx. \end{aligned} \tag{16}$$

For (16) to be meaningful, we must discuss the convergence of the improper integrals. Let us suppose $2 \leq a < b$. We then have $-\lambda a + c \leq -a\log 2 + \log\frac{5}{2} \leq \log\frac{5}{8} < 0$. Therefore, it follows that

$$\begin{aligned} &\int_a^b 1-\left(1-e^{-\lambda t+c}\right)^n dt = \int_{1-e^{-\lambda a+c}}^{1-e^{-\lambda b+c}} \frac{1-u^n}{\lambda(1-u)} du = \frac{1}{\lambda}\sum_{l=0}^{n-1} \int_{1-e^{-\lambda a+c}}^{1-e^{-\lambda b+c}} u^l du \\ &= \frac{1}{\lambda}\sum_{l=1}^{n} \frac{(1-e^{-\lambda b+c})^l - (1-e^{-\lambda a+c})^l}{l} \overset{b\to\infty}{\to} \frac{H_n}{\lambda} - \frac{1}{\lambda}\sum_{l=1}^{n} \frac{(1-e^{-\lambda a+c})^l}{l}, \end{aligned}$$

where $H_n = \sum_{k=1}^{n} \frac{1}{k}$. From this calculation, we see that the improper integrals in (16) exist. Note that the finite sum is bounded by

$$0 < 1-e^{-\lambda a+c} < \sum_{l=1}^{n} \frac{(1-e^{-\lambda a+c})^l}{l} < -\log\left\{1-(1-e^{-\lambda a+c})\right\} = \lambda a - c.$$

Hence, when $a = 2$ and $a = 3$, it turns out that

$$\int_2^\infty 1 - \left(1 - e^{-\lambda t + c}\right)^n dt \le \frac{H_n}{\lambda} - \frac{1 - e^{-2\lambda + c}}{\lambda} = \frac{H_n}{-\log(2r)} - \frac{1 - 10r^2}{-\log(2r)}$$

and

$$\int_3^\infty 1 - \left(1 - e^{-\lambda t + c}\right)^n dt \ge \frac{H_n}{\lambda} - \frac{3\lambda - c}{\lambda} = \frac{H_n}{-\log(2r)} + \frac{\log(2.5)}{-\log(2r)} - 3.$$

By applying them to (16) and the inequality $\log n < H_n < \log n + \gamma$, we obtain (15).

3.1. *Concrete bounds for* $\mathrm{E}(X_{n,n})$

If (15) is used, then $\mathrm{E}(X_{n:n})$ is bounded by

$$\begin{aligned} &6 - (1 - 4r + 2r^2)^n - (1 - 10r^2)^n + 2\left\{\frac{\log n}{-\log(2r)} + \frac{\log(2.5)}{-\log(2r)} - 3\right\} \\ &\le \mathrm{E}(X_{n:n}) \le 6 - (1 - 4r + 2r^2)^n - (1 - 10r^2)^n + 2\left\{\frac{\log n}{-\log(2r)} - \frac{1 - 10r^2 - \gamma}{-\log(2r)}\right\}. \end{aligned}$$

In particular, when $r = \frac{1}{4}$, i.e., $p = \frac{1}{2}$, it follows that

$$\begin{aligned} &2\log_2 n - \left(\frac{1}{8}\right)^n - \left(\frac{3}{8}\right)^n + 2\log_2 \frac{5}{2} \\ &\le \mathrm{E}(X_{n:n}) \le 2\log_2 n - \left(\frac{1}{8}\right)^n - \left(\frac{3}{8}\right)^n + 6 + \left(2\gamma - \frac{3}{4}\right)\log_2 e, \end{aligned}$$

more specifically, we have

$$\begin{aligned} 9.2 &\le \mathrm{E}(X_{10:10}) &&\le 13.3, \\ 15.9 &\le \mathrm{E}(X_{100:100}) &&\le 19.9, \\ 22.5 &\le \mathrm{E}(X_{1000:1000}) &&\le 26.6, \\ 29.2 &\le \mathrm{E}(X_{10000:10000}) &&\le 33.2. \end{aligned}$$

Under this idealized model, the expected maximum length $\mathrm{E}(X_{n,n})$ remains relatively modest, even when $p = \frac{1}{2}$, due to the light-tailed nature of the distribution of X (see (7)).

Acknowledgements

This research was supported by KAKENHI JP19K03622 and JP24K06847 of Japan Society for the Promotion of Science.

References

1. Cooper, C., Kennedy, R. (2021). Expected length and probability of winning a tennis game. *Math. Gaz.* 105(564): 490-500. doi:10.1017/mag.2021.117

2. Liu, Y. (Fall 2001). Random walks in tennis, *Missouri J. Math. Sci.* 13(3): 154-162. doi: 10.35834/2001/1303154

3. Stewart, I. (1989). *Game, set and math: Enigmas and conundrums*, Basil Blackwell, Oxford.

4. Grimmett, G., Stirzaker, D. (2020). *Probability and random processes*, 4th ed., Oxford UP, Oxford.

TOSHIO NAKATA

Department of Mathematics, University of Teacher Education Fukuoka,
Munakata, Fukuoka, 811-4192, Japan
e-mail: *nakata@fukuoka-edu.ac.jp*