Ann Inst Stat Math
DOI 10.1007/s10463-008-0214-0

Gambler's ruin and winning a series by *m* games

Tamás Lengyel

Received: 27 February 2008 / Revised: 4 September 2008

Abstract Two teams play a series of games until one team accumulates m more wins than the other. These series are fairly common in some sports provided that the competition has already extended beyond some number of games. We generalize these schemes to allow ties in the single games. Different approaches offer different advantages in calculating the winning probabilities and the distribution of the duration N, including difference equations, conditioning, explicit and implicit path counting, generating functions and a martingale-based derivation of the probability and moment generating functions of N. The main result of the paper is the determination of the exact distribution of N for a series of fair games without ties as a sum of independent geometrically distributed random variables and its approximation.

Keywords Gambler's ruin · Distribution of the duration · Martingales · Probability and moment generating functions · Limit theorem · Chebyshev polynomial of the first kind

1 Introduction

Two teams play a series of games until one team accumulates m more wins than the other. Each game has three possible outcomes: team (or player as used interchangeably in this paper) A wins with probability p, B wins with probability q ($0 < p,\ q < 1$), or they tie with probability $r = 1 - p - q \geq 0$. The series ends when one team has won m more games than the other and thus becomes the winner of the series. From now on $P(A)$ and $P(B) = 1 - P(A)$ denote the respective probabilities that teams A and B win the series, and N is the number of games played in the series or the duration.

T. Lengyel (✉)
Occidental College, 1600 Campus Road, Los Angeles, CA 90041, USA
e-mail: lengyel@oxy.edu

Published online: 22 January 2009

The purpose of any series is to magnify the differences between the teams. Tennis is discussed in Kemeny and Snell (1960) (in which "points" and "game" take on the role of game and series of games, respectively) as a hybrid between two popular procedures. To win a "game" in tennis the winner is required to have four wins (as in the World Series or a "4 best of 7" series) and be ahead by two. Apparently, Bernoulli was the first to analyze tennis. He found a difference-equation-based double recurrence for the winning probability $P(A)$ and showed that $P(A)$ can be written as a ratio of two seventh degree polynomials in p/q (cf. Blom et al. 1994). He also proposed handicapping in favor of the weaker player in order to balance $P(A)$ and $P(B)$.

If $p = q = 1/2$ then the World Series with $m^2/2$ required wins and the winning by m games (or win-by-m games) series are comparable in the sense that they yield an approximate mean duration of m^2 for large values of m, cf. Feller (1968), Kemeny and Snell (1960), Lengyel (1993), and Menon and Indira (1983). (For the asymptotic magnitude of the variance see Menon and Indira (1983).) In the general case, the former magnifies minute differences in $p - q$ by about $0.8m$, while the latter multiplies them by m, thus making the latter series more efficient and favorable. Any mixture, e.g., a "game" in tennis, lies in between in terms of efficiency. In fact, Siegrist (1989) studied special hybrids, the (n, k) contests in which the first team or player to win at least n games and to be ahead of its opponent by at least k games wins the contest. Championship series are often in a (4, 1) format (e.g., World Series), a tennis "game" is in a (4, 2) format while a tennis set (without tiebreaker) is in a (6, 2) format. Our win-by-m games problem is simply a contest in the (m, m) format. Siegrist obtained results for the probability of winning and the expected length of the (n, k) contest, and compared different formats from the point of view of the duration and the power of these contests as "tests" in order to determine the stronger team or player. For example, he observed that (3, 2) is a better format than (4, 1) in both senses if p and q are sufficiently close to 0.5.

Gambler's ruin problems offer a special case of the win-by-m games series. In fact, in this paper we will only consider this case. Assume that each of two players has a capital of m dollars. In each game a dollar can change hands between the two players: player A pays a dollar to player B with probability p or a dollar is paid to player A by player B with probability q, and no money is exchanged with probability $r = 1 - p - q$. The game is over when one player goes bankrupt, i.e., when the other player amasses m more wins.

In some cases, and typically when we want to determine the winning probabilities only, we can ignore games that end in a tie, and therefore, we will then use

$$p' = \frac{p}{p+q} \quad \text{and} \quad q' = \frac{q}{p+q} \tag{1}$$

to denote the probability of winning and losing a game, respectively, given that the game is not a tie.

In Sect. 2, we give a brief historical overview. Various random walk based approaches are presented in Sect. 3. They lead us to different derivations of the winning probabilities but not necessarily of the expected duration $E(N)$.

The final Sect. 4 is devoted to martingales, and it shows a fairly simple way to derive the probability generating function $p_m(x, p, q)$ and the moments of N. We determine the exact distribution of N and some of its asymptotic properties for $p = q = 1/2$ in Theorems 5 and 6, by using its moment generating function. We prove that N can be viewed as a sum of independent but not identically distributed random variables of various geometric distributions. The main results of the paper are summarized in Theorems 3, 5, and 6. Some details will be left to the reader.

A generalization of the gambler's ruin problem to higher dimensions is considered in Kmet and Petkovšek (2002). The exact and asymptotic expected duration is determined in some special cases with identical goals in each dimension. It corresponds to playing a series of different types of games and stopping when a player wins by m games in any type.

Interested readers can find other ways of generalizing gambler's ruin in Flajolet and Huillet (2008). They discuss an urn-based model which can be applied to a modification of the gambler's ruin in which the single game winning probabilities are affected by the number of previous wins and losses, e.g., the winning probability increases as the accumulated number of wins does. This modification results in a decreased expected duration. The authors exhibit limit theorems and a decomposition of the duration into a sum of independent random variables of different geometric distributions in the case corresponding to winning by m games.

2 Gambler's ruin by difference equations and other approaches

Gambler's ruin problems are typically represented by a random walk on the set of integers. A point moves to the right or to the left with probability p or q, or stays in place with probability $r = 1 - p - q$. The walk ends when it hits either of the two absorbing states m and $-m$. To make referencing easier we set the independent and identically distributed random variables

$$Z_i = \begin{cases} 1, & \text{with probability } p, \\ -1, & \text{with probability } q, \\ 0, & \text{with probability } r, \end{cases} \tag{2}$$

$i = 1, 2, \ldots, n$, indicating win or loss by team A, or tie in the ith game. Clearly, $S_n = \sum_{i=1}^{n} Z_i$ is the difference in the number of games won by teams A and B after n games, and team A wins the series if for some $n : S_n = m$ and $|S_k| \leq m - 1$ for all $k < n$. The determination of $P(A)$, the probability of this occurring, is referred to as Huygens' fifth problem in Blom et al. (1994).

Often recurrence relations, implied by difference equations, are used to determine both the winning probabilities and the expected length of the game as in Feller (1968, Chapter XIV) and Problem 1582 (1999). For example, it can be seen in Feller (1968, volume 1, identity (2.4) on p. 345) that

$$P(A) = \frac{\lambda^m}{1 + \lambda^m} = \frac{p^m}{p^m + q^m} \quad \text{and} \quad P(B) = \frac{q^m}{p^m + q^m} \tag{3}$$

with

$$\lambda = p'/q' = p/q \tag{4}$$

if $p \neq q$, and $P(A) = P(B) = 1/2$ otherwise. The role of λ will become transparent in the martingale approach (Sect. 4). Note that the winning probabilities depend only on the p to q ratio and not on r (as ties will only affect the length of the game). In addition to this,

$$E(N) = \begin{cases} \dfrac{m(\lambda^m - 1)}{(p-q)(\lambda^m + 1)} = \dfrac{m(P(A) - P(B))}{p - q}, & \text{if } p \neq q, \\ \dfrac{m^2}{2p}, & \text{if } p = q, \end{cases} \tag{5}$$

cf. Feller (1968, volume 1, Problem #5, p. 367). Some related problems and facts are discussed in Lengyel (2009). It also offers a way to introduce ties to the classical problem via a decomposition of the duration N into a random sum of N' independent and identically distributed random variables, the ith term being an arbitrary nonnegative number T_i, $i = 1, 2, \ldots, N'$, of ties immediately before the ith win or loss followed by the win or loss, i.e., $N = \sum_{i=1}^{N'}(T_i + 1)$. The distribution of $T_i + 1$ is geometric with parameter $1 - r$. Thus, for instance, we can relate $E(N)$ to $E(N')$, the expected duration in the classical problem, by Wald's identity.

An alternative approach for determining $P(A)$ and $E(N)$ is to apply the theory of Markov chains (cf. Kemeny and Snell 1960). We can also use conditioning to calculate $P(A)$. One way to find probabilities of "competing" events is to use a conditional setting. We consider the random walk (with no absorbing states) on the set of integers that starts at 0. Now, let E, E_1 and E_2 be the events that the random walk ever visits $-m$, visits m before $-m$, and $-m$ before m, respectively. It is easy to see (e.g., Feller 1968, volume 1, identity (2.8) on p. 347) that

$$P(E) = \begin{cases} \left(\dfrac{q}{p}\right)^m, & \text{if } p > q, \\ 1, & \text{if } p \leq q. \end{cases} \tag{6}$$

For $p > q$, we have $(\frac{q}{p})^m = P(E) = P(E|E_1)P(E_1) + P(E|E_2)P(E_2) = (\frac{q}{p})^{2m} P(A) + (1 - P(A))$ which immediately implies (3). Note that identity (6) can be rephrased as $P(\sup_{n \geq 0} S_n \geq m) = (p/q)^m$ if $q > p$ and $m \geq 0$, thus identifying the distribution of $\sup_{n \geq 0} S_n$ as geometric.

3 Lattice path counting and generating functions

We can use implicit and explicit lattice path counting and generating functions to derive the winning probabilities and the distribution of duration.

3.1 Path counting

The first approach requires only implicit calculations. The probability $P(A)$ is

$$P(A) = \sum_{\substack{n \geq 0 \\ t \geq 0}} c(n, m, t)\, p^{n+m} q^n r^t \tag{7}$$

where the factor $c(n, m, t)$ counts the number of ways we can arrange $n+m$ wins, n losses, and t ties (from the point of view of team A) so that team A reaches its goal of being ahead of team B by exactly m games for the first time after the last game. By symmetry,

$$P(B) = \sum_{\substack{n \geq 0 \\ t \geq 0}} c(n, m, t)\, q^{n+m} p^n r^t.$$

We observe that $P(A)/P(B) = p^m/q^m$ and $P(A) + P(B) = 1$; therefore,

$$P(A) = \frac{p^m}{p^m + q^m} \quad \text{and} \quad P(B) = \frac{q^m}{p^m + q^m}.$$

On the other hand, $c(n, m, t)$ explicitly counts the number of ways a random walk on the plane going from $(0, 0)$ to $(2n + m + t, m)$ (i.e., with $n + m$ up steps $(1, 1)$, n down steps $(1, -1)$, and t horizontal steps $(1, 0)$) first reaches the boundary $|y| = m$ on its last move. We define $d(n, m) = c(n, m, 0)$. Clearly,

$$c(n, m, t) = \binom{2n + m + t - 1}{t} d(n, m) \tag{8}$$

for the last move cannot be a $(1, 0)$.

For $m = 2$ we have $c(n, 2, t) = \binom{2n+2+t-1}{t} 2^n$ since after pairing the non-horizontal moves, each pair contains a $(1, 1)$ and a $(1, -1)$ move which yields $d(n, 2) = 2^n$ (cf. Problem 1582 (1999)). If $m = 3$ then a recurrence-based approach (cf. Stern 1979) or a standard block walking argument yields

Theorem 1 *For $m = 2$ and 3, the number of paths on the plane from $(0, 0)$ to $(2n + m, m)$ with $n + m$ up steps $(1, 1)$ and n down steps $(1, -1)$ and first passage to $|y| = m$ on the last move is $d(n, m) = m^n$.*

Now, by plugging this into (7), we can determine $P(A)$. Alternatively, we can ignore the tied games and focus on winning and losing the other games, but now with corresponding probabilities p' and q' given in (1). Clearly, for $m = 2$ and 3, $P(A) = \sum_{n=0}^{\infty} d(n, m)(p')^{n+m}(q')^n = \sum_{n=0}^{\infty} m^n (p')^{n+m}(q')^n = (p')^m/(1 - mp'q')$ $= p^m/(p^m + q^m)$, and the expected value and standard deviation of the duration can be easily computed by (8). For example, if $p = q = 1/2$ then $(E(N), \sigma(N)) = (4, 2\sqrt{2})$ and $(9, 4\sqrt{3})$ if $m = 2$ and 3, respectively.

In general, we can use the theory of lattice path counting of Mohanty (1979) and Narayana (1979), Theorem 2 of Sect. 2.2 of Mohanty (1979) in particular, and that of the enumeration of Dyck paths to obtain the number of paths going between two (horizontal) boundaries, but the calculations become cumbersome beyond small values of m.

3.2 Generating functions

By using the generating function of the probability of absorption at m (i.e., winning by team A) at the nth game (Feller 1968, volume 1, pp. 349–351), we can systematically, though implicitly, obtain $d(n, m)$ and any moment of the random variable N. In fact, if $r = 0$ then, by using difference equations (Feller 1968, Chapter XIV) derives

$$g_m(x, p, q) = \sum_{n=0}^{\infty} P(N = n, S_n = m)\, x^n = \frac{1}{\lambda_1^m(x) + \lambda_2^m(x)} \tag{9}$$

and for the probability generating function of the duration N of the game

$$\begin{aligned} p_m(x, p, q) &= \sum_{n=0}^{\infty} P(N = n)\, x^n \\ &= \left(\frac{1}{\lambda^m} + 1\right) \frac{1}{\lambda_1^m(x) + \lambda_2^m(x)} = \frac{1}{P(A)} g_m(x, p, q) \end{aligned} \tag{10}$$

with $\lambda_1(x) = \frac{1+\sqrt{1-4pqx^2}}{2px}$ and $\lambda_2(x) = \frac{1-\sqrt{1-4pqx^2}}{2px}$. Remarkably, the length of the game has no effect on the winning probabilities (cf. Samuels 1975).

Theorem 2 (Samuels) *The duration N and the end point S_N, i.e., who wins, are independent random variables.*

Interested readers can find a proof using generating functions and an extension to the case with ties allowed in single games in Lengyel (2009). (From now on Theorem 2 refers to the extended version.) For a general $r \geq 0$, by calculations similar to that in Feller (1968), we can prove

Theorem 3 *We set $r = 1 - p - q \geq 0$, $\lambda = p/q$, $\lambda_1(x) = \frac{1-rx+\sqrt{(1-rx)^2-4pqx^2}}{2px}$, and $\lambda_2(x) = \frac{1-rx-\sqrt{(1-rx)^2-4pqx^2}}{2px}$. The generating function of the probability of the duration with team A winning at the nth game and the probability generating function of the duration N of the game are given by* (9) *and* (10)*, respectively. The duration N has no effect on the winning probabilities.*

We note that the independence also follows by a simple argument similar to the one used in Sect. 3.1.

Clearly, $g_1(x, p, q) = \frac{px}{1-rx}$ and $g_m(x, p, q) = \frac{p^m x^m}{(1-rx)^{m-2}((1-rx)^2-mpqx^2)}$ for $m = 2$ and 3 by Sect. 3.1. We note that $P(A) = g_m(1, p, q)$ easily reduces to (3) and $E(N) = g_m'(1, p, q) + g_m'(1, q, p) = p_m'(1, p, q)$ leads to (5) since it is easy to see that the power series $g_m(x, p, q)$ and $p_m(x, p, q)$ are both convergent in an open circle of radius $1/(1 - (\sqrt{p} - \sqrt{q})^2)$ which thus contains 1 if $p \neq q$. After intensive simplifications, we can determine $\mathrm{var}(N) = p_m''(1, p, q) + p_m'(1, p, q) - (p_m'(1, p, q))^2$. If $p = q$ then by Abel's convergence theorem, the remaining part of (5) and $\mathrm{var}(N) = \frac{m^2(2m^2+1-6p)}{12p^2}$ follow. This latter yields $\mathrm{var}(N) = 4m\binom{m+1}{3}$ if $p = q = 1/2$ (as it was observed for $m = 2$ and 3 in Sect. 3.1). We note that $g''_m(1, p, q)$ was determined for $p = q = 1/2$ in Beyer and Waterman (1979), and Bach (1997) obtained the first six moments and cumulants of N. He also discussed the arithmetic complexity of computing the rth moment and a connection to Brownian motion. Aoyama et al. (2008) use a similar technique to determine the exact first-passage time distribution of a modified random walk. We will present a martingale-based alternative proof of Theorem 3 in Sect. 4.3 (although one relying on Theorem 2).

We note that Kac (1945) obtained the exact probability of the duration in the form of an alternating trigonometric sum of m terms for $p = q = 1/2$. This guarantees an asymptotically exponential decrease of $P(N = n)$ at the rate $\cos(\pi/2m)$ as $n \to \infty$. Karni attempted to find a simple form for $P(N = n)$ if $p + q = 1$ in Karni (1977, 1978), but only succeeded with some restriction on the length of the duration.

4 Martingale approach

In this section we approach our problems by defining associated martingales (cf. Baldi et al. 2002; Blom et al. 1994; Lalley 2003; Williams 1991). We say that the sequence $\{Y_n\}$ forms a martingale with respect to the sequence of random variables $\{X_n\}$ if $E(|Y_n|) < \infty$ and $E(Y_n|X_1, X_2, \ldots, X_{n-1}) = Y_{n-1}, n \geq 2$. We observe the X_is until they satisfy some prescribed stopping condition. We call the number N of the observed X_is a stopping time. If $P(N < \infty) = 1$, $E(|Y_N|) < \infty$, and $\lim_{n\to\infty} E(Y_n|N > n)P(N > n) = 0$ then $E(Y_N) = E(Y_1)$ by the Optional Stopping Theorem (e.g., Blom et al. 1994).

Note that for our stopping rule $E(N) < \infty$ and $\lim_{n\to\infty} P(N > n) = 0$. In fact, by a standard argument, the series can be viewed in blocks of $2m$ consecutive games. If a block corresponds to a run (or winning streak) of $2m$ wins for either team then that team wins (unless the game has already ended). Therefore, $P(N > 2m) \leq 1 - p^{2m}$ and similarly, $P(N > k \cdot 2m) \leq (1 - p^{2m})^k$, and the distribution of N exhibits an exponentially decaying right tail.

We note that for any $i \geq 1$, $M_n = S_n - nE(Z_i) = Z_1 + Z_2 + \cdots + Z_n - nE(Z_i)$ defines a martingale with respect to the independent and identically distributed random variables Z_is given in (2). In general, Wald's (first) equation (e.g., Baldi et al. 2002; Blom et al. 1994; Lalley 2003) yields $E(S_N) = E(N)E(Z_i)$ since $E(N) < \infty$.

We also define the constant λ in order to guarantee $E(\lambda^{-Z_i}) = 1$. This yields $\lambda = p/q$, in agreement with (4), independently of r. It can be verified that the sequence

$$R_n = \lambda^{-S_n}, n = 1, 2, \ldots, \tag{11}$$

is also a martingale with respect to the $Z_i s$. (Sometimes it is referred to as Wald's martingale.) In fact, $E(R_n|Z_1, Z_2, \ldots, Z_{n-1}) = \lambda^{-S_{n-1}} E(\lambda^{-Z_n}) = \lambda^{-S_{n-1}} = R_{n-1}$.

4.1 Winning probability

Note that $E(|R_N|)$ and $E(R_n|N > n)$ both are bounded from above by $\max\{\lambda^{-m}, \lambda^m\}$, thus the Optional Stopping Theorem applies: $E(R_N) = E(R_1) = E(\lambda^{-Z_1}) = 1$. On the other hand,

$$E(R_N) = E(\lambda^{-S_N}) = P(A)\lambda^{-m} + P(B)\lambda^m = \lambda^m + P(A)(\lambda^{-m} - \lambda^m)$$

which yields (3) for $p \neq q$ (cf. Feller 1968).

4.2 Expected length

If $\lambda \neq 1$ then by Wald's equation and $E(Z_i) = p - q \neq 0$ it follows that

$$E(N) = \frac{E(S_N)}{E(Z_i)} = \frac{mP(A) - mP(B)}{p - q},$$

and by (3)

$$E(N) = \frac{m(\lambda^m - 1)}{(p - q)(\lambda^m + 1)}.$$

If $p = q$ then we can use Wald's second equation (e.g., Baldi et al. 2002, p. 37): $E\left((S_N - NE(Z_i))^2\right) = E(N)\,\mathrm{var}(Z_i)$ which turns into $E(N) = E(S_N^2)/\mathrm{var}(Z_i) = m^2/(2p)$, verifying the remaining part of (5).

4.3 Higher moments of N

The random variable R_n is a special case of likelihood ratio martingales (see Feller 1968, volume 2, pp. 211–212; Lalley 2003). In general,

$$R_n(\theta) = \prod_{i=1}^{n} \frac{e^{\theta Z_i}}{\phi(\theta)}$$

with the moment generating function $\phi(\theta) = E(e^{\theta Z_i}) = pe^{\theta} + qe^{-\theta} + r$ of Z_i for any real θ. As above, by the Optional Stopping Theorem, we get $E(R_N(\theta)) = E(R_1(\theta)) = 1$. Now we can observe that $\phi(-\ln\lambda) = 1$ and thus, in fact, the R_n defined in (11) is $R_n = R_n(-\ln\lambda) = \lambda^{-S_n}$.

Remark Let $p_{Z_i}(s) = \sum_k s^k P(Z_i = k)$ be the probability generating function of Z_i (this time defined for an integer valued random variable Z_i taking both positive and negative values; therefore, $p_{Z_i}(s)$ is a Laurent polynomial). If $\lambda \neq 1$ then the equation $p_{Z_i}(s) = E(s^{Z_i}) = \phi(\ln s) = 1$ has two roots: $s = 1/\lambda = q/p$ and the trivial $s = 1$. We note that this and other properties of $p_{Z_i}(s)$ are applied in Ethier and Khoshnevisan (2002) to obtain bounds on $P(A)$ for a more complicated profit variable Z_i.

The convex function $\phi(\theta)$ takes its minimum $1-(\sqrt{p}-\sqrt{q})^2$ at $\theta_{\min} = -\frac{1}{2}\ln\lambda$. Let us assume that $\theta \geq 0$ and $\lambda = p/q > 1$ which guarantee that $\phi(\theta) \geq 1$. This will also be the case for all θ if $p = q$. Then, by way of Wald's third equation in Lalley (2003), i.e., for any bounded stopping time $N \wedge n$ (the truncation to the smaller of N and n) : $1 = E(R_{N\wedge n}(\theta)) = E\left(\frac{e^{\theta S_{N\wedge n}}}{\phi^{N\wedge n}(\theta)}\right)$, and by the dominated convergence theorem as

$n \to \infty$, it follows that

$$1 = E\left(\frac{e^{\theta S_N}}{\phi^N(\theta)}\right). \tag{12}$$

In fact, $\lim_{n\to\infty} \frac{e^{\theta S_{N\wedge n}}}{\phi^{N\wedge n}(\theta)} = \frac{e^{\theta S_N}}{\phi^N(\theta)}$ since we have $P(N < \infty) = 1$ (as $E(N)$ is finite here), $e^{\theta S_{N\wedge n}} \le \max\{e^{\theta m}, e^{-\theta m}\}$ and $\phi(\theta) \ge 1$; thus, $\frac{e^{\theta S_{N\wedge n}}}{\phi^{N\wedge n}(\theta)} \le \max\{e^{\theta m}, e^{-\theta m}\}$ for all $n \ge 0$.

By an argument similar to the derivation of the probability generating function of the first passage time to 1 in Williams (1991), the probability generating function (10) can be also easily derived from (12) without the technical overhead of difference equations referred to in Sect. 3.2. Toward this end, we now substitute $1/x = \phi(\theta) = p/u + qu + r$ with $u = e^{-\theta}$, which yields $u = \frac{1-rx-\sqrt{(1-rx)^2-4pqx^2}}{2qx}$ and $u^{-1} = \frac{1-rx+\sqrt{(1-rx)^2-4pqx^2}}{2px}$ since $x < 1$ implies $\theta > 0$ and thus $u < 1$ if $p > q$. If $p = q$ then $u \le u^{-1}$ for all $x \le 1$, thus $u \le 1$. We shall need Theorem 2 of Sect. 3.2. By conditioning in (12) we obtain that $1 = \sum_{n=0}^{\infty} E(e^{\theta S_N} x^N \mid N = n, S_n = m)\, P(N = n, S_n = m) + \sum_{n=0}^{\infty} E(e^{\theta S_N} x^N \mid N = n, S_n = -m)\, P(N = n, S_n = -m) = \sum_{n=m}^{\infty} e^{\theta m} P(A)\, P(N = n)\, x^n + \sum_{n=m}^{\infty} e^{-\theta m} P(B)\, P(N = n)\, x^n = E(e^{\theta S_N})E(x^N)$. Thus,

$$p_N(x) = E(x^N) = 1/E(e^{\theta S_N}) = 1/(P(A)u^{-m} + P(B)u^m) \tag{13}$$

and (10) follows immediately for all $x \le 1$ (since $1/x = \phi(\theta) \ge 1$) and arbitrary choices of p and q (as the case $q > p$ is similar). Note that we can extend the range of x if $p \ne q$ by Sect. 3.2.

In addition to this, if $p \ne q$ then by expanding $p_N(x)$ about $x = 1$ we can determine the moments of N. We define the inverse function of ϕ: let $\phi^{-1}(u)$ be the unique value $v : -\frac{1}{2}\ln\lambda \le v < \infty$, so that $\phi(u) = v$. (The case $\lambda < 1$ is similar.) Note that we can take the derivative of $\phi^{-1}(u)$ repeatedly around 1. From $p_N(x) = 1/E(e^{\theta S_N})$ with $\theta = \phi^{-1}(1/x)$ we can derive the probability generating function of N which can help us to determine exactly or to approximate the probability distribution of N. In fact, carrying out the moment calculations by applying the approximation is a little easier than using (10) directly. For instance, by $p'_N(1) = E(N)$ and the first order approximation of the function ϕ^{-1} we get (5). The second order approximation and $\mathrm{var}(N) = p''_N(1) + p'_N(1) - (p'_N(1))^2$ lead us to the calculation of the variance of N provided that $p_N(x)$, thus $\phi^{-1}(x)$, is differentiable around 1, i.e., if $p \ne q$.

On the other hand, if $p = q$ then we can use our findings of Sect. 3.2. Combining these cases, we obtain

Theorem 4

$$\mathrm{var}(N) = \begin{cases} \dfrac{m(P(A) - P(B))}{p-q}\left(\dfrac{p+q}{(p-q)^2} - 1\right) - \dfrac{4m^2 P(A)P(B)}{(p-q)^2}, & \text{if } p \ne q \\ \dfrac{m^2(2m^2 + 1 - 6p)}{12p^2}, & \text{if } p = q \le 1/2. \end{cases}$$

Higher moments of N can be derived in a similar fashion. The moment generating function of N can be obtained by $M_N(t) = E(e^{tN}) = p_N(e^t)$ (or simply by substituting $t = -\ln\phi(\theta)$ into identity $E(1/\phi^N(\theta)) = 1/E(e^{\theta S_N})$ if $p \neq q$).

4.4 The distribution of N if $p = q = 1/2$

Note that for the contest in $(m, 1)$ format, i.e., the best m of $2m-1$ (or World) series, the convergence of the duration WS_m to the normal variable is completely characterized in Menon and Indira (1983). Unfortunately, as they note it, the normal approximation is not valid for values of p close to 0, 1/2, or 1. In the case of $p = q = 1/2$, Stadje (1998) found that $(2m - WS_m)/(\sqrt{2m})$, as $m \to \infty$, has the limit distribution of the absolute value of a standard normal random variable. We also note that a remarkable closed form was given by Stirzaker (1988) for the double generating function of $2m - WS_m$, if $p = q = 1/2$.

We now focus on the distribution of the duration N for contests in (m, m) format with $p = q = 1/2$. The moment generating function of Z_i is $\phi(\theta) = \cosh(\theta)$, hence by identity (13) the probability generating and moment generating functions of N are $p_N(x) = 1/\cosh(m\theta) = 1/T_m(\cosh(\theta)) = 1/T_m(1/x)$ with $\theta = \operatorname{arccosh}(1/x) \geq 0$ (cf. Feller 1968, volume 1, identity (5.4) on p. 352) and

$$M_N(t) = \frac{1}{T_m(e^{-t})}, \tag{14}$$

respectively, with $T_m(x)$ being the mth Chebyshev polynomial of the first kind (cf. Comtet 1974; Weisstein 2002). As one of the referees pointed it out, this relation can be also derived directly from (9) using the fact that

$$T_m(x) = \frac{(x - \sqrt{x^2-1})^m}{2} + \frac{(x + \sqrt{x^2-1})^m}{2}.$$

Note that the relation of Chebyshev polynomials of the second kind to Dyck path enumeration has been explored elsewhere, e.g., in Krattenthaler (2001).

In this case, we find below the exact distribution of N in terms of a sum of independent but not identically distributed random variables of various geometric distributions (Theorem 5). We also derive the limit Theorem 6 which can be used to approximate the distribution of N. We note that the methods presented here and leading to Theorems 5 and 6 do not seem to generalize to unequal single game winning probabilities or when $r > 0$.

4.4.1 *Exact distribution*

Theorem 5 *Let $p = q = 1/2$ and $r_i = \cos\frac{(2i-1)\pi}{2m}$, $i = 1, 2, \ldots, m$, be the roots of the mth Chebyshev polynomial of the first kind. We define $R_i : 0 < R_i = -r_i r_{m-i+1} = r_i^2 < 1, i = 1, 2, \ldots, \lfloor m/2 \rfloor$, and consider $\lfloor m/2 \rfloor$ independent random variables $X_i \sim \text{Geometric}(1 - R_i), i = 1, 2, \ldots, \lfloor m/2 \rfloor$. In this case, the distribution of the* *half-duration $N/2$ is identical to that of $\sum_{i=1}^{\lfloor m/2\rfloor} X_i + \delta_m$ with $\delta_m = 1/2$ if m is odd and 0 otherwise, yielding $E(N) = m^2$ and $var(N) = 2m^2(m^2-1)/3$.*

Proof of Theorem 5. We factor the mth Chebyshev polynomial to find a decomposition of random variable $N/2$ into a sum of other variables. Observe that $T_m(x) = 2^{m-1}\prod_{i=1}^{m}(x - \cos\frac{(2i-1)\pi}{2m})$ (cf. Weisstein 2002). Clearly, the roots $\cos\frac{(2i-1)\pi}{2m}$ and $\cos\frac{(2m-2i+1)\pi}{2m}$, $i = 1, 2, \ldots, \lfloor m/2\rfloor$, are symmetric about zero. Thus every $R_i = -r_i r_{m-i+1}$, $i = 1, 2, \ldots, \lfloor m/2\rfloor$, falls strictly between zero and one. If m is odd then there is an extra root at zero.

We prove the theorem for m even. In this case, we have $T_m(x) = 2^{m-1}\prod_{i=1}^{\lfloor m/2\rfloor}(x^2 - R_i)$. First we note that

$$\prod_{i=1}^{m}(1-r_i) = \prod_{i=1}^{\lfloor m/2\rfloor}(1-R_i) = 1/2^{m-1} \tag{15}$$

by the generating function (cf. Comtet 1974; Weisstein 2002)

$$g(t,x) = \frac{1-xt}{1-2xt+t^2} = \sum_{m=0}^{\infty} T_m(x)t^m = \frac{1}{2} + \frac{1}{2}\frac{1-t^2}{1-2xt+t^2}. \tag{16}$$

In fact, we have $T_m(1) = 2^{m-1}\prod_{i=1}^{\lfloor m/2\rfloor}(1-R_i) = 2^{m-1}\prod_{i=1}^{m}(1-r_i)$. On the other hand, the coefficient of t^m of $g(t,1) = 1/(1-t)$ is equal to $T_m(1)$, hence $[t^m]\, g(t,1) = 1 = 2^{m-1}\prod_{i=1}^{m}(1-r_i)$.

Therefore, by identities (14) and (15) we have that $M_{N/2}(t) = \frac{1}{T_m(e^{-t/2})} = \frac{1}{2^{m-1}\prod_{i=1}^{\lfloor m/2\rfloor}(e^{-t}-R_i)} = \frac{1}{2^{m-1}\prod_{i=1}^{\lfloor m/2\rfloor}(1-R_i)}\prod_{i=1}^{\lfloor m/2\rfloor}\frac{(1-R_i)e^t}{1-R_i e^t} = \prod_{i=1}^{\lfloor m/2\rfloor} M_{X_i}(t)$ with $X_i \sim$ Geometric$(1-R_i)$, $i = 1, 2, \ldots, \lfloor m/2\rfloor$.

If m is odd then there is an extra factor x in $T_m(x)$ which results in an extra factor $e^{t/2}$ in $M_N(t/2)$ which is the moment generating function of the constant 1/2. We leave the details to the reader. The expected value and variance of N follow by Sect. 4.2 and Theorem 4. □

Remark An alternative derivation of $E(N)$ and var(N) also follows from the above decomposition, and higher moments can be computed similarly. First, we define the generating function of the kth power sum of the roots of $T_m(x)$ by $S(x) = \sum_{k=1}^{\infty} S_k x^k$, $S_k = \sum_{i=1}^{m} r_i^k$, $k = 1, 2, \ldots$, and the corresponding alternating generating function of the elementary symmetric polynomials: $\Pi(x) = \sum_{k=0}^{\infty}(-1)^k \Pi_k x^k$, $\Pi_0 = 1$, $\Pi_k = \sum_{1\le i_1<i_2<\cdots<i_k\le m} r_{i_1}r_{i_2}\cdots r_{i_k}$, $k = 1, 2\ldots, m$, and $\Pi_k = 0$ if $k > m$. The Newton–Girard formulas can be rewritten as

$$-x\Pi'(x)/\Pi(x) = S(x). \tag{17}$$

Next we derive that $\mu_m = \sum_{i=1}^{\lfloor m/2\rfloor}\frac{1}{1-R_i} + \delta_m = m^2/2$ and $\sigma_m^2 = \sum_{i=1}^{\lfloor m/2\rfloor}\frac{R_i}{(1-R_i)^2} = m^2(m^2-1)/6$ for the half-duration. To prove these identities we observe that

$\sum_{i=1}^m \frac{1}{1-r_i} = 2\big(\sum_{i=1}^{\lfloor m/2\rfloor} \frac{1}{1-R_i} + \delta_m\big)$ and $\sum_{i=1}^m \frac{r_i}{(1-r_i)^2} = 4\sum_{i=1}^{\lfloor m/2\rfloor} \frac{R_i}{(1-R_i)^2}$, then develop the left hand sides as series involving power sums of the roots r_i. For instance, $\sum_{i=1}^m \frac{1}{1-r_i} = \sum_{i=1}^m \sum_{k=0}^\infty r_i^k = \sum_{k=0}^\infty \sum_{i=1}^m r_i^k = \sum_{k=0}^\infty S_k$ and similarly, $\sum_{i=1}^m \frac{r_i}{(1-r_i)^2} = \sum_{k=1}^\infty kS_k$.

We specialize (17) by setting $\Pi(x) = \prod_{i=1}^m (1 - r_i x) = x^m T_m(1/x)/2^{m-1}$ which yields

$$S(x) = -m + \frac{T_m'(1/x)}{xT_m(1/x)}. \tag{18}$$

Clearly, $T_m(1) = 1$, $T_m'(1) = m^2$, $T_m''(1) = m^2(m^2-1)/3$, and $T_m'''(1) = 8m\binom{m+2}{5}$ by deriving the partial derivatives $g_x(t,1)$, $g_{xx}(t,1)$, and $g_{xxx}(t,1)$ based on (16). For instance, $\sum_{m=1}^\infty T_m'(1)t^m = g_x(t,1) = \frac{t(1+t)}{(1-t)^3} = \sum_{m=1}^\infty m^2 t^m$. In general, using a standard formula to calculate the nth derivative of the reciprocal of the function $1-2xt+t^2$, we can derive that $T_m^{(n)}(1) = 2^{n-1}n!\big(\binom{m+n}{m-n} + \binom{m+n-1}{m-n-1}\big)$ for $n \ge 1$. Now we can calculate $S(1)+m = \sum_{k=0}^\infty S_k$ and $S'(1) = \sum_{k=1}^\infty kS_k$ by (18) to obtain μ_m and σ_m^2.

Despite the decomposition, other moments require more involved calculations, e.g., we need $\sum_{i=1}^{\lfloor m/2\rfloor} \frac{2-3R_i+R_i^2}{(1-R_i)^3}$ in order to obtain the third central moment of the half-duration.

We also note that there has been some interest in determining the asymptotic behavior of the raw moments of the duration N as $m \to \infty$. The normalized duration N/m^2 captures the asymptotic features even better. In fact, the first few raw moments are $\mu_1' = 1$, $\mu_2' = 5/3 + o(1)$, $\mu_3' = 61/15 + o(1)$, $\mu_4' = 277/21 + o(1)$ and $\mu_5' = 50521/945 + o(1)$ as $m \to \infty$, and according to Bach (1997), the error terms are functions of $1/m^2$. The moment $\mu_k' = E\big((N/m^2)^k\big)$ converges to a finite positive constant c_k for any integer $k \ge 1$ as $m \to \infty$, however, determining c_k for large values of k remains a numerically challenging problem.

4.4.2 Approximating the distribution

Unfortunately, none of the usual criteria (cf. Weisstein 2002), e.g., the Lyapunov condition, work here and thus, this approach does not guarantee the normal limit law for N. However, we can approximate the distribution of N by using only a small percentage of the largest terms of the decomposition in Theorem 5. For instance, let $f(m) = \lfloor cm \rfloor$ with any $c : 0 < c < 1$, and take the sum $Y_1 = \sum_{i=1}^{f(m)} X_i$. The above arguments show that $N/2 = Y_1 + E_1$ so that $E(2Y_1) \sim m^2$, $E(E_1) = \Theta(m)$, $\text{var}(2Y_1) \sim 2m^4/3$, and $\text{var}(E_1) = \Theta(m)$.

Taking this a little further, we can approximate N by using only an asymptotically zero percent of the terms plus an approximately normally distributed error term of a smaller magnitude.

Theorem 6 *Let $f(m) = \lfloor m^{1-\epsilon} \rfloor$ with any $\epsilon : 0 < \epsilon < 1/5$ and consider the sum $Y_2 = \sum_{i=1}^{f(m)} X_i$ (where the X_is are as in Theorem 5). For $N/2 = Y_2 + E_2$ we get that*

$E(2Y_2) \sim m^2$, $E(E_2) = O(m^{1+2\epsilon})$, $\mathrm{var}(2Y_2) \sim 2m^4/3$, and $\mathrm{var}(E_2) = O(m^{1+4\epsilon})$ and E_2 has an asymptotically normal distribution.

Proof of Theorem 6. We consider only the case with m even. In fact, we need that $\sum_{i=1}^{\lfloor m/2\rfloor} \frac{1}{1-R_i} = \sum_{i=1}^{f(m)} \frac{1}{((2i-1)\pi/(2m))^2} + mO(\frac{m^2}{f(m)^2}) = \frac{4m^2}{\pi^2}\sum_{i=1}^{f(m)} \frac{1}{(2i-1)^2} + O(m^{1+2\epsilon}) = \frac{4m^2}{\pi^2}\frac{\pi^2}{6}\frac{3}{4} + O(m^{1+2\epsilon}) = \frac{m^2}{2} + O(m^{1+2\epsilon})$. Here we used the approximation $1 - R_i = \left(\frac{(2i-1)\pi}{2m}\right)^2 + O(m^{-4})$ for $R_i = \cos^2\frac{(2i-1)\pi}{2m}$ where $i \le cm$ for some sufficiently small $c > 0$. Similarly, $\sum_{i=1}^{\lfloor m/2\rfloor}\frac{R_i}{(1-R_i)^2} = \sum_{i=1}^{f(m)}\frac{1}{((2i-1)\pi/(2m))^4} - \sum_{i=1}^{f(m)}\frac{1}{((2i-1)\pi/(2m))^2} + mO(\frac{m^4}{f(m)^4}) = \frac{16m^4}{\pi^4}\sum_{i=1}^{f(m)}\frac{1}{(2i-1)^4} - \frac{m^2}{2} + O(m^{1+4\epsilon}) = \frac{16m^4}{\pi^4}\frac{\pi^4}{90}\frac{15}{16} - \frac{m^2}{2} + O(m^{1+4\epsilon}) = \frac{m^4}{6} - \frac{m^2}{2} + O(m^{1+4\epsilon})$.

We set $g(x,n) = \sum_{k=0}^{\infty} k^n x^k = \frac{A_n(x)}{(1-x)^{n+1}}, n \ge 1$, with $A_n(x)$ being the nth Eulerian polynomial (cf. Comtet 1974; Lengyel 1996), which guarantees that $g(1-p,k) \le k!p^{-(k+1)}$ since $A_k(x) \le k!$ if $|x| \le 1$. Observe that for the raw moments $\mu'_k = E(X_i^k) = \sum_{j=1}^{\infty} j^k(1-p)^{j-1}p = \frac{p}{1-p}g(1-p,k) \le \frac{p}{1-p}k!p^{-(k+1)} = O(m^{2k}/f(m)^{2k})$ holds with $p = 1 - R_i = \Theta(f(m)^2/m^2)$ for $i : f(m) \le i \le (1+\delta)f(m)$ with any sufficiently small $\delta > 0$. This implies an upper bound on the magnitude of the central moments $\mu_k = E((X_i - EX_i)^k) = \sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\mu'_j(\mu'_1)^{k-j} = O(m^{2k}/f(m)^{2k})$ by induction on k.

Now we check the Lyapunov condition (Weisstein 2002) for $E_2 = \sum_{i=f(m)+1}^{\lfloor m/2\rfloor} X_i$ with some positive α. First we note that $E\left((X_i - EX_i)^{2+\alpha}\right) \le P(|X_i - EX_i| < 1) + E\left((X_i - EX_i)^4\right) = E\left((X_i - EX_i)^4\right) + O(1)$ as $m \to \infty$. This leads to

$$\begin{aligned}
&\frac{\max_{f(m)+1\le i\le\lfloor m/2\rfloor} E(|X_i - EX_i|^{2+\alpha})}{\sum_{i=f(m)+1}^{\lfloor m/2\rfloor} E(|X_i - EX_i|^{2+\alpha})} \\
&\le \frac{\max_{f(m)+1\le i\le\lfloor m/2\rfloor} E(|X_i - EX_i|^4) + a_m}{\sum_{i=f(m)+1}^{\lfloor m/2\rfloor} E((X_i - EX_i)^2) - b_m} \\
&\le \frac{E\left((X_{f(m)+1} - EX_{f(m)+1})^4\right) + a_m}{\sum_{i=f(m)+1}^{\lfloor m/2\rfloor} \mathrm{var}(X_i) - b_m} \\
&\le \frac{\frac{Cm^8}{f(m)^8}}{\sum_{i=f(m)+1}^{\lfloor m/2\rfloor}\frac{R_i}{(1-R_i)^2}} \le \frac{\frac{Cm^8}{f(m)^8}}{\sum_{i=f(m)+1}^{\lfloor(1+\delta)f(m)\rfloor}\frac{R_i}{(1-R_i)^2}} \\
&= O\left(\frac{\frac{m^8}{f(m)^8}}{f(m)\frac{m^4}{f(m)^4}}\right) = O\left(\frac{m^4}{f(m)^5}\right) = o(1)
\end{aligned}$$

for any sufficiently large $C > 0$, positive $a_m = O(1)$ and $b_m = O(m)$ as $m \to \infty$. Hence the limit of the leftmost ratio is 0, and the approximate normality of E_2 follows. □

From the above proof it is clear that the condition on $\epsilon : 0 < \epsilon < 1/5$ can be improved.

Acknowledgments The author wishes to thank the two referees and Gregory P. Tollisen for raising some open questions and making helpful suggestions and comments that improved the presentation of this paper.

References

Aoyama, K., Shimizu, K., Ong, S. H. (2008). A first-passage time random walk distribution with five transition probabilities: a generalization of the shifted inverse trinomial. *Annals of the Institute of Statistical Mathematics 60*, 1–20.

Bach, E. (1997). Moments in the duration of play. *Statistics & Probability Letters 36*, 1–7.

Baldi, P., Mazliak, L., Priouret, P. (2002). *Martingales and Markov Chains, Solved Exercises and Elements of Theory*. Boca Raton, FL: Chapman & Hall/CRC.

Beyer, W. A., Waterman, M. S. (1979). Moments of absorption time for a conditioned random walk. *Studies in Applied Mathematics 60*, 83–90.

Blom, G., Holst, L., Sandell, D. (1994). *Problems and snapshots from the world of probability*. New York–Berlin–Heidelberg: Springer.

Comtet, L. (1974). *Advanced combinatorics*. Dordrecht: D. Reidel.

Ethier, S. N., Khoshnevisan, D. (2002). Bounds on gambler's ruin probabilities in terms of moments. *Methodology and Computing in Applied Probability 4*, 55–68.

Feller, W. (1968). *Introduction to probability theory and its applications*, volumes 1-2. Wiley, New York, 3rd edition.

Flajolet, P., Huillet, T. (2008). Analytic combinatorics of the mabinogion urn. http://algo.inria.fr/flajolet/Publications/FlHu08.pdf, Submitted to the Fifth Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities, September 22–26, 2008, Blaubeuren, Germany.

Kac, M. (1945). Random walk in the presence of absorbing barriers. *Annals of Mathematical Statistics 16*, 62–67.

Karni, E. (1977). The probability distribution of the duration of the game in the classical ruin problem. *Journal of Applied Probability 14*, 416–420.

Karni, E. (1978). Correction: The probability distribution of the duration of the game in the classical ruin problem. *Journal of Applied Probability 15*, 865–865.

Kemeny, G., Snell, J. L. (1960). *Finite Markov Chains*. The University Series in Undergraduate Mathematics, Van Nostrand: Princeton, NJ.

Kmet, A., Petkovšek, M. (2002). Gambler's ruin problem in several dimensions. *Advances in Applied Mathematics 28*, 107–118.

Krattenthaler, C. (2001). Permutations with restricted patterns and Dyck paths. *Advances in Applied Mathematics*, 27:510–530. Special issue in honor of Dominique Foata's 65th birthday (Philadelphia, PA, 2000).

Lalley, S. (2003). Conditional expectation and martingales. http://galton.uchicago.edu/~lalley/Courses/313/Martingales.pdf.

Lengyel, T. (1993). A combinatorial identity and the World Series. *SIAM Review 35*, 294–297.

Lengyel, T. (1996). On some properties of the series $\sum_{k\geq 0} k^n x^k$ and the Stirling number. *Discrete Mathematics 150*, 281–292.

Lengyel, T. (2009). The conditional gambler's ruin problem with ties allowed. *Applied Mathematics Letters 22*, 351–355.

Menon, V. V., Indira, N. K. (1983). On the asymptotic normality of the number of replications of a paired comparison. *Journal of Applied Probability 20*, 554–562.

Mohanty, S. G. (1979). *Lattice Paths Counting and Applications*. New York: Academic Press.

Narayana, T. V. (1979). *Lattice Paths Combinatorics with Statistical Applications*. Toronto: University of Toronto Press.

Problem 1582 (1999). Problem 1582. *Mathematics Magazine*, 72: Solution in 73(2000), October issue, 326–327.

Samuels, S. M. (1975). The classical ruin problem with equal initial fortunes. *Mathematics Magazine 48*, 286–288.

Siegrist, K. (1989). n-point, win-by-k games. *Journal of Applied Probability 26*, 807–814.

Stadje, W. (1998). Asymptotic probabilities in a sequential urn scheme related to the matchbox problem. *Advances in Applied Probability 30*, 831–849.

Stern, F. (1979). Absorption sequences. *Fibonacci Quarterly 17*, 275–280.

Stirzaker, D. (1988). A generalization of the matchbox problem. *The Mathematical Scientist 13*, 104–114.
Weisstein, E. W. (2002). *CRC concise encyclopedia of mathematics* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
Williams, D. (1991). *Probability with martingales*. Cambridge: Cambridge University Press.