

## AVERAGE EXTREMA OF A RANDOM WALK WITH A NEGATIVE BINOMIAL STOPPING TIME

David K. Neal

Department of Mathematics  
Western Kentucky University  
Bowling Green, KY 42101, USA

**Abstract:** A random walk  $X$  moves upward or downward one unit at a time with probabilities  $p$  and  $q = 1 - p$ , respectively. We derive the averages of the maximum height and the minimum height attained before  $n$  downward movements occur, and show the relationship between these average extrema and the average final height. We derive the limits of these average extrema as  $n$  increases to  $\infty$ , and derive the condition on  $p$  that makes the average extrema symmetric about the initial height of  $X$ .

**AMS Subject Classification:** 60G50

**Key Words:** simple random walk, average maximum height, negative binomial random variable, reflection principle

### 1. Introduction

Random walks provide a rich structure to study a variety of mathematical concepts, including combinatorics, graph theory, calculus, probability, and analysis. The upward and downward movements alone produce the binary form of the interval  $[0, 1]$ . Adding further structure and conditions allow for a variety of applications including modeling gambling fortunes for progressive betting strategies (see [5]).

In this paper we shall consider a one-dimensional random walk that is stopped after a fixed number of downward movements. For each number of upward steps that occur, a family of symmetric graphs is created that we shall analyze to determine various properties such as the number of ways each particular height can be attained. We then shall use this analysis to determine the average maximum height and the average minimum height that are attained before stopping and to show the relationship between these averages and the average final height. We also derive the limits of these average extrema as the number of downward movements increases, and derive a surprisingly simple condition that makes the average extrema symmetric about the initial height of the process.

## 2. The Stopped Process

Throughout, we let  $(X, j, p)$  be the height of a simple random walk that begins at integer height  $j$  and moves either up or down one unit at a time on each independent step with probabilities  $p$  and  $q = 1 - p$ , where  $p \neq 1$ . Steps are denoted on the  $x$ -axis so that a path moves linearly from coordinates  $(0, j)$  to either  $(1, j + 1)$  or  $(1, j - 1)$  on the first step. We let  $M_i$  and  $m_i$  be the maximum and minimum heights attained through  $i$  steps, and we allow the walk to continue until exactly  $n$  downward movements have occurred. Then the number of steps  $T_n$  taken by the various possible paths is a negative binomial random variable:  $T_n \sim nb(n, q)$ .

Although only  $n$  downward steps occur before stopping, paths may take an unlimited number of upward steps; thus, the maximum height  $M_{T_n}$  of a stopped path may be any integer greater than or equal to  $j$ . However for each fixed number  $k$  of upward movements that actually occur, there are only a finite number of paths that are possible, and therefore only a finite number of maximum heights attained by these paths. By partitioning these paths into the numbers of ways each maximum height can occur, we can obtain the conditional average maximum height of paths that have taken exactly  $k$  upward steps before the final downward step. These conditional averages allow us to find the average maximum height  $E[M_{T_n}]$  attained through  $T_n$  steps. A similar counting technique allows us to find the average minimum height  $E[m_{(T_n-1)}]$  attained before the last step, as well as the overall average minimum height  $E[m_{T_n}]$ .

The random walk  $X$  stops on step  $T_n$  after  $n$  downward movements have occurred. Suppose that there have been exactly  $k$  upward movements before

stopping so that a total of  $n + k$  steps have been taken. Because the last step must be downward, the  $k$  upward movements must occur within the first  $n + k - 1$  steps. Thus, the number of ways for a path to stop in  $n + k$  steps, for  $k \geq 0$ , is given by

$$S_{n,k} = \binom{n+k-1}{k}. \tag{1}$$

The pdf of the negative binomial random variable  $T_n$  gives the probability of a path stopping with exactly  $k$  upward steps and is given by

$$P(T_n = n + k) = S_{n,k} p^k q^n, \tag{2}$$

for  $k \geq 0$ , with the average number of steps being  $E[T_n] = n/q$ . Using this distribution, we can find the average number of upward steps taken before stopping, the average final height upon stopping, and an upper bound for the average maximum height.

**Theorem 1.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements on step  $T_n$ . Let  $U_{T_n}$  be the number of upward steps taken before stopping, and let  $M_{T_n}$  and  $m_{T_n}$  be the maximum and minimum values attained through step  $T_n$ . Then*

- (a)  $E[U_{T_n}] = np/q$ ,
- (b)  $E[X_{T_n}] = j - n + np/q$ ,
- (c)  $j - n \leq E[m_{T_n}] \leq E[m_{(T_n-1)}] \leq j \leq E[M_{(T_n-1)}] = E[M_{T_n}] \leq j + np/q$ .

*Proof.* (a) Upon stopping, there have been  $k$  upward steps if and only if  $T_n = n + k$ ; thus,

$$\begin{aligned} E[U_{T_n}] &= \sum_{k=0}^{\infty} k P(T_n = n + k) = \sum_{k=0}^{\infty} (n + k - n) P(T_n = n + k) \\ &= \sum_{k=0}^{\infty} (n + k) P(T_n = n + k) - n = E[T_n] - n \\ &= \frac{n}{q} - n = \frac{n(1 - q)}{q} = \frac{np}{q}. \end{aligned}$$

(b) Because  $X$  moves up or down one unit at a time, the final height of a stopped path is  $X_{T_n} = j - n + U_{T_n}$ ; thus, (b) follows from (a).

(c) Because  $X$  only takes  $n$  downward steps of one unit each, the height cannot drop below  $j - n$ . Also, an additional step can only lower the minimum

value of a path, and the minimum values cannot be more than the initial height of  $j$ . Thus,  $j - n \leq E[m_{T_n}] \leq E[m_{(T_n-1)}] \leq j$ . The final downward step cannot change the maximum height attained through  $T_n - 1$  steps and maximum values must be at least the initial height of  $j$ ; therefore, we always have  $j \leq M_{(T_n-1)} = M_{T_n}$ . And the maximum value can be no more than  $j$  plus the number of upward steps taken; that is,  $M_{T_n} \leq j + U_{T_n}$ . Hence,  $j \leq E[M_{(T_n-1)}] = E[M_{T_n}] \leq j + E[U_{T_n}] = j + np/q$ .  $\square$

Because the last step is always downward, the height  $X_{(T_n-1)}$  before the last step will always be one unit higher than  $X_{T_n}$ . Thus, we may state:

**Corollary 1.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements on step  $T_n$ . The average height of  $X$  on the next to last step is  $E[X_{(T_n-1)}] = j - n + np/q + 1$ .*

### 3. Conditional Mean Maximum Height

In order to find the average maximum height through step  $T_n$ , we first will find the conditional average maximum height of paths that have made  $k$  upward movements and a total of  $n + k$  steps. Because the last step must be downward, the height after  $n + k - 1$  steps is  $j + k - n + 1$ . So  $S_{n,k}$  actually gives the total number of ways for a path to move from coordinates  $(0, j)$  to  $(n + k - 1, j + k - n + 1)$  while moving up or down one unit at a time as steps are taken along the  $x$ -axis.

Among these  $S_{n,k}$  paths, we can determine the maximum heights that may occur and count the number of paths that attain those heights. For the single path that goes upward  $k$  times in a row then downward  $n$  times, the maximum height attained is  $j + k$ , which is the largest possible maximum value. Before the last downward step, the height must be  $j + k - n + 1$ . So if  $k \geq n$ , then paths must reach a height greater than the initial height  $j$ . But if  $k < n$ , then it is possible for paths to initially decrease and never surpass the initial height. Therefore the possible maximum heights are of the form  $j + i$ , for  $i = \max(0, k - n + 1), \dots, k$ . Thus, there are  $k + 1$  possible maximum heights if  $k < n$ , and  $n$  possible maximum heights if  $k \geq n$ .

Figure 1 shows all possible paths that begin at height  $j = 2$ , end after  $n = 4$  downward steps, and make a total of  $k = 7$  upward steps. There are  $S_{4,7} = \binom{10}{7} = 120$  such paths. Here all paths must reach a height of at least  $j + k - n + 1 = 6$ , and the possible maximum heights are of the form  $2 + i$ , for  $i = 4, \dots, 7$ .

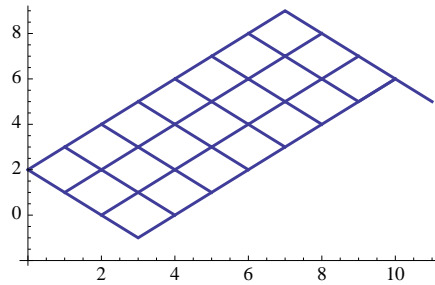


Figure 1: 7 upward steps before 4 downward

Now we shall count the number of paths  $U_{n,k,i}$  that have a maximum height of  $j + i$ , for  $i = \max(0, k - n + 1), \dots, k$ . However by subtracting the initial height  $j$  from the height of each path,  $S_{n,k}$  also gives the number of ways for a path to move from coordinates  $(0, 0)$  to  $(n + k - 1, k - n + 1)$ , while  $U_{n,k,i}$  now gives the number of paths that have a maximum height of  $i$ , for  $i = \max(0, k - n + 1), \dots, k$ .

By the *Reflection Principle* (see [2]), the number of paths that begin at  $(0, 0)$ , touch or surpass the integer height  $i \geq 0$ , and then end at coordinates  $(N, K)$  is given by

$$B_{N,K,i} = \binom{N}{\frac{N+2i-K}{2}}.$$

Then the number of paths that start at  $(0, 0)$ , end at  $(N, K)$ , and have a *maximum* value of  $i$  is given by

$$B_{N,K,i} - B_{N,K,(i+1)} = \binom{N}{\frac{N+2i-K}{2}} - \binom{N}{\frac{N+2i+2-K}{2}}.$$

Now letting  $N = n + k - 1$  and  $K = k - n + 1$ , we find that the number of paths with  $k$  upward movements that have a maximum height of  $j + i$  is

$$U_{n,k,i} = \binom{n+k-1}{n+i-1} - \binom{n+k-1}{n+i}, \tag{3}$$

for  $i = \max(0, k - n + 1), \dots, k - 1$ , with  $U_{n,k,k} = \binom{n+k-1}{n+k-1} = 1$ .

One can easily verify that the telescoping values  $U_{n,k,i}$  always sum to  $S_{n,k}$ . Now because each individual path is equally likely, the probability of obtaining a maximum height of  $j + i$  is then  $U_{n,k,i}/S_{n,k}$ , which allows us to compute the average maximum height of these paths. We state the result as follows:

**Lemma 1.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements on step  $T_n$ . Among paths that make exactly  $k$  upward movements, the average maximum height is*

$$E[M_{T_n} \mid T_n = n + k] = j + \sum_{i = \max(0, k - n + 1)}^k \frac{i U_{n,k,i}}{S_{n,k}}.$$

**Example 1.** Consider the paths in Figure 1 with initial height  $j = 2$ ,  $k = 7$  upward steps, and  $n = 4$  downward steps. Of the  $S_{4,7} = \binom{10}{7} = 120$  possible paths, there are  $\binom{10}{7} - \binom{10}{8} = 75$  paths with a maximum height of 6,  $\binom{10}{8} - \binom{10}{9} = 35$  paths with a maximum height of 7,  $\binom{10}{9} - \binom{10}{10} = 9$  paths with a maximum height of 8, and  $\binom{10}{10} = 1$  path with a maximum height of 9. The conditional average maximum height is then  $E[M_{T_4} \mid T_4 = 11] = 97/15 = 6.4\bar{6}$ .

#### 4. Conditional Mean Minimum Height

We also can determine the average minimum height among paths that have exactly  $k$  upward movements before stopping. If  $k = 0$ , then clearly the average minimum height is  $j - n$ . But if  $k \geq 1$ , then the range of  $m_{T_n}$  depends on if  $k \geq n$  or if  $1 \leq k < n$ . In either case, the lowest height that can be attained is  $j - n + 1$ . But after all  $n + k$  steps, the final height must still be  $j + k - n$ . So if  $k \geq n$ , then the last downward step cannot reduce the minimum height, and a minimum height of  $j$  or lower must be attained before the final step. In other words,  $m_{T_n} = m_{(T_n-1)}$  when  $k \geq n$ .

But if  $k < n$ , then  $X$  must drop below the initial height  $j$  at some point and the minimum height may not be attained until the final downward step. Thus for  $1 \leq k < n$ , we have  $m_{T_n} \leq m_{(T_n-1)}$  and  $m_{T_n} \leq j - (n - k)$ . In general, for  $k \geq 1$ , the minimum height  $m_{T_n}$  attained by a path before stopping must be of the form  $j - h$  for  $h = \max(0, n - k), \dots, n - 1$ .

Figure 2 shows all possible paths that begin at height  $j = 2$ , end after  $n = 7$  downward steps, and make a total of  $k = 5$  upward steps. There are  $S_{7,5} = \binom{11}{5} = 462$  such paths. The final height is 0 and it is possible for the minimum height to be attained only on the last step.

For simplification, we first shall analyze  $m_{(T_n-1)}$ . Then for  $1 \leq k < n$ , we can make a simple adjustment to account for  $m_{T_n}$  possibly occurring on the last step. After  $T_n - 1$  steps, the final height is  $j + k - n + 1$ ; thus,  $j - (n - 1) \leq m_{(T_n-1)} \leq j - (n - k - 1)$ . That is, for  $k \geq 1$ ,  $m_{(T_n-1)}$  must be of the form  $j - h$  for  $h = \max(0, n - k - 1), \dots, n - 1$ . So  $m_{(T_n-1)}$  has  $k + 1$  possible values if  $1 \leq k < n$ , and  $n$  possible values if  $k \geq n$ .

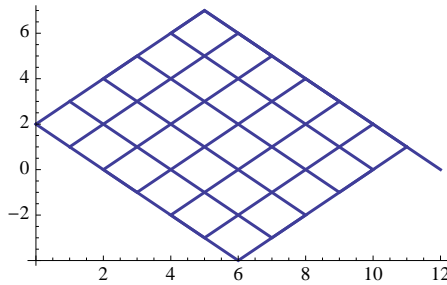


Figure 2: 5 upward steps before 7 downward

We now shall count the number of paths  $L_{n,k,h}$  for which  $m_{(T_n-1)}$  has a minimum value of  $j - h$ . But, by symmetry,  $L_{n,k,h}$  will equal  $U_{n,k,i}$  for a particular  $i$ . Indeed, when  $1 \leq k < n$ , then the number of paths with minimum values of  $j - h$ , for each  $h$  from  $n - k - 1$  to  $n - 1$ , equals the number of paths with maximum values of  $j + i$  for  $i$  from  $0$  to  $k$ , where  $i = h - n + k + 1$ . And when  $k \geq n$ , then the number of paths with minimum values of  $j - h$ , for each  $h$  from  $0$  to  $n - 1$ , equals the number of paths with maximum values of  $j + i$  for  $i$  from  $k - n + 1$  to  $k$ , again where  $i = h - n + k + 1$ .

Substituting for  $i$  in (3), we obtain the number of paths with  $k$  upward movements for which  $m_{(T_n-1)}$  has a minimum height of  $j - h$  to be

$$L_{n,k,h} = U_{n,k,(h-n+k+1)} = \binom{n+k-1}{k+h} - \binom{n+k-1}{k+h+1}, \tag{4}$$

for  $h = \max(0, n - k - 1), \dots, n - 2$ , with  $L_{n,k,n-1} = \binom{n+k-1}{n+k-1} = 1$ .

Using Formula (4), we can express the conditional average minimum height before the last step as follows:

**Lemma 2.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements on step  $T_n$ . Among paths that have exactly  $k$  upward movements, the average minimum height before the last downward step is*

$$E[m_{(T_n-1)} | T_n = n + k] = j - \sum_{h = \max(0, n-k-1)}^{n-1} \frac{h L_{n,k,h}}{S_{n,k}}.$$

We note that the formula in Lemma 2 does hold for  $k = 0$  and gives  $E[m_{(T_n-1)} | T_n = n] = j - n + 1$ .

We now can adjust the result to obtain the conditional average minimum height through the last step. When  $k \geq n$ , then the last downward step cannot

change the minimum height. But for  $k < n$ , there are  $L_{n,k,(n-k-1)}$  paths with  $m_{(T_n-1)} = j - n + k + 1$  which then attain a final minimum height of  $m_{T_n} = j - n + k$  on the last step. Thus in the conditional average in Lemma 2, we must subtract 1 unit with probability  $L_{n,k,(n-k-1)}/S_{n,k} = (n - k)/n$ . Thus we obtain:

**Corollary 2.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements on step  $T_n$ . Among paths that have exactly  $k$  upward movements, the average minimum height is*

$$E[m_{T_n} \mid T_n = n + k] = j - \sum_{h = \max(0, n-k-1)}^{n-1} \frac{h L_{n,k,h}}{S_{n,k}} - \max\left(0, \frac{n - k}{n}\right).$$

We note that for  $k = 0$  Corollary 2 gives  $E[m_{T_n} \mid T_n = n] = j - n$ .

**Example 2.** For the 120 paths in Figure 1, the last downward step cannot decrease the minimum height; thus, we have  $E[m_{T_4} \mid T_4 = 11] = E[m_{(T_4-1)} \mid T_4 = 11]$ . There are  $\binom{10}{7} - \binom{10}{8} = 75$  paths with  $m_{T_4} = 2$ ,  $\binom{10}{8} - \binom{10}{9} = 35$  paths with  $m_{T_4} = 1$ ,  $\binom{10}{9} - \binom{10}{10} = 9$  paths with  $m_{T_4} = 0$ , and 1 path with  $m_{T_4} = -1$ . Thus,  $E[m_{(T_4-1)} \mid T_4 = 11] = E[m_{T_4} \mid T_4 = 11] = 23/15 = 1.5\bar{3}$ .

**Example 3.** For the 462 paths in Figure 2 with  $j = 2$ ,  $k = 5$ , and  $n = 7$ ,  $m_{(T_7-1)}$  is of the form  $2 - h$  for  $1 \leq h \leq 6$ . There are  $\binom{11}{6} - \binom{11}{7} = 132$  paths with  $m_{(T_7-1)} = 1$ ,  $\binom{11}{7} - \binom{11}{8} = 165$  paths with  $m_{(T_7-1)} = 0$ ,  $\binom{11}{8} - \binom{11}{9} = 110$  paths with  $m_{(T_7-1)} = -1$ ,  $\binom{11}{9} - \binom{11}{10} = 44$  paths with  $m_{(T_7-1)} = -2$ ,  $\binom{11}{10} - \binom{11}{11} = 10$  paths with  $m_{(T_7-1)} = -3$ , and  $\binom{11}{11} = 1$  path with  $m_{(T_7-1)} = -4$ . We then have  $E[m_{(T_7-1)} \mid T_7 = 12] = -50/231 \approx -0.21645$  and  $E[m_{T_7} \mid T_7 = 12] = -50/231 - 2/7 = -116/231 \approx -0.5021645$ .

The possible maximum heights are of the form  $2 + i$ , for  $i = 0, \dots, 5$ . By symmetry, there are 132 paths with  $M_{T_7} = 2$ , 165 paths with  $M_{T_7} = 3$ , 110 paths with  $M_{T_7} = 4$ , 44 paths with  $M_{T_7} = 5$ , 10 paths with  $M_{T_7} = 6$ , and 1 path with  $M_{T_7} = 7$ . Thus,  $E[M_{T_7} \mid T_7 = 12] = 743/231$ .

Using the symmetry of paths through  $T_n - 1$  steps, we obtain a simple formula for the sum of the conditional extreme values:

**Lemma 3.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements at time  $T_n$ . For  $k \geq 0$ , the sum of the conditional extreme values satisfies*

$$E[M_{T_n} + m_{(T_n-1)} \mid T_n = n + k] = 2j - n + k + 1.$$



*Proof.* Among the  $S_{n,k}$  paths that have exactly  $k$  upward steps, consider a fixed  $h$  from  $\max(0, n - k - 1)$  to  $n - 1$ . Then there are an equal number of paths for which  $m_{(T_n-1)} = j - h$  as there are paths for which  $M_{T_n} = j + i$ , where  $i = h - n + k + 1$ . So for these  $L_{n,k,h}$  paths, the sum of the extreme values is  $2j - n + k + 1$ . Because all paths are equally likely and all collections of  $L_{n,k,h}$  paths have a common sum of extreme values, the result follows.  $\square$

**Example 4.** From Examples 1 and 2, the sum of the average extrema for the paths in Figure 1 is  $E[M_{T_4} + m_{(T_4-1)} \mid T_4 = 11] = 97/15 + 23/15 = 8$ , which is the value given by Lemma 3 with  $j = 2$ ,  $n = 4$ , and  $k = 7$ . For the paths in Figure 2, the results from Example 3 give  $E[M_{T_7} + m_{(T_7-1)} \mid T_7 = 12] = 743/231 - 50/231 = 3$ , which is the value given by Lemma 3 with  $j = 2$ ,  $n = 7$ , and  $k = 5$ .

### 5. The Average Extreme Heights

The conditional averages given in Lemmas 1, 2, and 3 do not depend on the probabilities  $p$  and  $q$  of  $X$  moving upward and downward. However the total expected values do depend on these probabilities. By summing the weighted conditional averages, we obtain the average extreme values of the stopped random walk.

**Theorem 2.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements at time  $T_n$ .*

(a) *The average maximum height attained through the time of stopping is*

$$E[M_{T_n}] = j + q^n \sum_{k=0}^{\infty} \left( \sum_{i=\max(0, k-n+1)}^k i U_{n,k,i} \right) p^k,$$

where  $U_{n,k,i} = \binom{n+k-1}{n+i-1} - \binom{n+k-1}{n+i}$ , for  $i = \max(0, k - n + 1), \dots, k - 1$ , with  $U_{n,k,k} = 1$ .

(b) *The average minimum heights attained before the last step and through the time of stopping are*

$$E[m_{(T_n-1)}] = j - q^n \sum_{k=0}^{\infty} \left( \sum_{h=\max(0, n-k-1)}^{n-1} h L_{n,k,h} \right) p^k,$$

and

$$E[m_{T_n}] = E[m_{(T_n-1)}] - \frac{q^n}{n} \sum_{k=0}^{n-1} (n-k) \binom{n+k-1}{k} p^k,$$

where  $L_{n,k,h} = \binom{n+k-1}{k+h} - \binom{n+k-1}{k+h+1}$  for  $h = \max(0, n-k-1), \dots, n-2$ , with  $L_{n,k,n-1} = 1$ .

*Proof.* Using Lemma 1 and (2) we have

$$\begin{aligned} E[M_{T_n}] &= \sum_{k=0}^{\infty} E[M_{T_n} \mid T_n = n+k] \times P(T_n = n+k) \\ &= j + \sum_{k=0}^{\infty} \left( \sum_{i=\max(0, k-n+1)}^k \frac{i U_{n,k,i}}{S_{n,k}} \right) S_{n,k} p^k q^n, \end{aligned}$$

and (a) follows. The result for  $E[m_{(T_n-1)}]$  follows similarly using Lemma 2, and the result for  $E[m_{T_n}]$  follows from Lemma 3 and (1). □

The formulas for the average extrema in Theorem 2 do not significantly simplify, but we can find a simple formula for their sum.

**Theorem 3.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements at time  $T_n$ . The average extreme values satisfy*

$$E[M_{T_n}] + E[m_{(T_n-1)}] = 2j - n + np/q + 1.$$

*Proof.* By summing the weighted conditional averages given in Lemma 3 and using Theorem 1(a), we obtain

$$\begin{aligned} E[M_{T_n}] + E[m_{(T_n-1)}] &= \sum_{k=0}^{\infty} (2j - n + k + 1) P(T_n = n+k) \\ &= (2j - n + 1) + \sum_{k=0}^{\infty} k P(T_n = n+k) \\ &= 2j - n + 1 + E[U_{T_n}] = 2j - n + 1 + \frac{np}{q}. \end{aligned}$$

□

We now can combine Theorem 3 and Corollary 1 to obtain a relationship between the average extrema and the average final height:

**Corollary 3.** *Let  $(X, j, p)$  be a simple random walk with  $p \neq 1$  that stops upon making  $n$  downward movements at time  $T_n$ . The average extreme values satisfy  $E[M_{T_n}] + E[m_{(T_n-1)}] - j = E[X_{(T_n-1)}] = E[X_{T_n}] + 1$ .*

**Example 5.** Consider the process  $(X, 2, 0.9)$  that is stopped after  $n = 4$  downward steps. The average number of steps taken is  $E[T_4] = 4/0.1 = 40$ . From Theorem 1(c), we have  $E[M_{T_4}] \leq 38$ , and according to Theorem 3, we have  $E[M_{T_4}] + E[m_{(T_4-1)}] = 37$ . The actual values for the average extrema given in Theorem 2 must be estimated numerically, but these infinite series converge quickly. Summing the expressions in Theorem 2 to  $k = 2000$  terms, we obtain  $E[M_{T_4}] \approx 35.12342$ ,  $E[m_{(T_4-1)}] \approx 1.87658$ , and  $E[m_{T_4}] \approx 1.87544$ , which gives  $E[M_{T_4}] + E[m_{(T_4-1)}] \approx 37$ .

### 6. The Average Maximum Height of a Non-Stopped Walk

Theorem 1(c) shows that the upper bound for  $E[M_{T_n}]$  grows as the number  $n$  of allowed downward steps increases. However, for  $p < q$ ,  $E[M_{T_n}]$  will remain bounded and attain a limit which will be the average maximum height of paths that are allowed to continue indefinitely. To show this result, we shall analyze the random variable  $M_\infty = \sup_{i \geq 0} X_i$ .

As a consequence of the classic *Boundary Problem* for simple random walks, the probability of  $(X, j, p)$  ever reaching an upper boundary height of  $N$ , for an integer  $N \geq j$ , is given by

$$P(M_\infty \geq N) = \begin{cases} 1 & \text{if } p \geq q \\ \left(\frac{p}{q}\right)^{N-j} & \text{if } p < q. \end{cases}$$

So for  $p \geq q$ , paths of  $X$  will almost surely reach any and every height  $N > j$ ; thus,  $M_\infty$  is infinite-valued almost surely and  $E[M_\infty] = \infty$ . But for  $p < q$ , the pdf of  $M_\infty$  is given by

$$\begin{aligned} P(M_\infty = N) &= P(M_\infty \geq N) - P(M_\infty \geq N + 1) \\ &= \left(\frac{p}{q}\right)^{N-j} \left(1 - \frac{p}{q}\right), \end{aligned} \tag{5}$$

for  $N \geq j$ . One can easily sum the geometric series in (5) to see that  $P(M_\infty < \infty) = 1$  for  $p < q$ . The distribution of  $M_\infty$  is of interest by itself, (see [1]). In particular, the tail behavior and its rate of convergence have been studied in detail (see [3] and [6]). We are only concerned though with the expected value of  $M_\infty$  which we derive next.

**Theorem 4.** *Let  $(X, j, p)$  be a simple random walk. For  $p < q$ , the average maximum height of  $X$  is given by*

$$E[M_\infty] = j + \frac{p}{q - p}.$$

*Proof.* Using the fact that  $\sum_{i=0}^\infty i x^i = x/(1 - x)^2$  for  $-1 < x < 1$ , we have

$$\begin{aligned} E[M_\infty] &= \sum_{N=j}^\infty N \times P(M_\infty = N) = \left(1 - \frac{p}{q}\right) \sum_{N=j}^\infty N \left(\frac{p}{q}\right)^{N-j} \\ &= \left(1 - \frac{p}{q}\right) \sum_{N=0}^\infty (j + N) \left(\frac{p}{q}\right)^N \\ &= j \left(1 - \frac{p}{q}\right) \sum_{N=0}^\infty \left(\frac{p}{q}\right)^N + \left(1 - \frac{p}{q}\right) \sum_{N=0}^\infty N \left(\frac{p}{q}\right)^N \\ &= j + (1 - p/q) \times \frac{p/q}{(1 - p/q)^2} = j + \frac{p}{q - p}. \end{aligned}$$

□

Reversing the roles of upward and downward and  $p$  and  $q$ , we also obtain the average minimum height of  $X$  for  $p > q$ :

**Corollary 4.** *Let  $(X, j, p)$  be a simple random walk. For  $p > q$ , the average minimum height of  $X$  is given by*

$$E[m_\infty] = j - \frac{q}{p - q}.$$

We now let  $M_n = \max_{0 \leq i \leq n} X_i$ . Then  $M_n$  is an everywhere increasing function of  $n$  since the maximum value of a path can only increase with more steps. And clearly  $M_n$  converges everywhere to  $M_\infty$ . So by the Monotone Convergence Theorem, we have  $\lim_{n \rightarrow \infty} E[M_n] = E[M_\infty]$ .

Because it must take at least  $n$  steps to obtain  $n$  downward movements, we have  $n \leq T_n$ . Then because  $M_n$  is an increasing function of  $n$ , we have  $M_n \leq M_{T_n} \leq M_\infty$ . Hence we also must have  $\lim_{n \rightarrow \infty} E[M_{T_n}] = E[M_\infty]$ . An analogous argument shows that  $\lim_{n \rightarrow \infty} E[m_{T_n}] = E[m_\infty]$ . We state the results as follows:

**Theorem 5.** *Let  $(X, j, p)$  be a simple random walk that stops upon making  $n$  downward movements at time  $T_n$ . The limits of the average extrema*

are

$$(a) \lim_{n \rightarrow \infty} E[M_{T_n}] = \begin{cases} \infty & \text{if } p \geq q \\ j + \frac{p}{q-p} & \text{if } p < q \end{cases}$$

$$(b) \lim_{n \rightarrow \infty} E[m_{T_n}] = \begin{cases} -\infty & \text{if } p \leq q \\ j - \frac{q}{p-q} & \text{if } p > q. \end{cases}$$

We can verify the results by applying Theorem 3 with increasing values of  $n$ . For  $(X, 2, 0.4)$ , we have we have  $E[M_{T_{50}}] \approx 3.9775$ ,  $E[M_{T_{100}}] \approx 3.99857$ , and  $E[M_{T_{200}}] \approx 3.99999$ , which verifies that  $\lim_{n \rightarrow \infty} E[M_{T_n}] = 4 = E[M_\infty]$ . For  $(X, 2, 0.9)$ , we have  $\lim_{n \rightarrow \infty} E[m_{T_n}] = 1.875 = E[m_\infty]$ .

Our formulas can easily be adjusted for a random walk  $(\tilde{X}, j, p, c)$  that moves up and down  $c$  units at a time. The average final height is now  $E[\tilde{X}_{T_n}] = j - cn + cnp/q$ , and  $E[\tilde{X}_{(T_n-1)}] = j - cn + cnp/q + c$ . If  $\tilde{M}_{T_n}$  and  $\tilde{m}_{T_n}$  are the extrema of  $\tilde{X}$  through time  $T_n$ , then  $E[\tilde{M}_{T_n}] = j + c(E[M_{T_n}] - j)$ , where  $E[M_{T_n}]$  is as in Theorem 2(a), while  $E[\tilde{m}_{(T_n-1)}] = j + c(E[m_{(T_n-1)}] - j)$ , and  $E[\tilde{m}_{T_n}] = E[\tilde{m}_{(T_n-1)}] - c(E[m_{(T_n-1)}] - E[m_{T_n}])$ .

That is, by Theorem 2, the average extrema of  $(X, j, p)$  are of the form  $E[M_{T_n}] = j + A_1$ ,  $E[m_{(T_n-1)}] = j - A_2$ , and  $E[m_{T_n}] = j - A_3$ . Then  $E[\tilde{M}_{T_n}] = j + cA_1$ ,  $E[\tilde{m}_{(T_n-1)}] = j - cA_2$ , and  $E[\tilde{m}_{T_n}] = j - cA_3$ , where the  $A_i$  are the requisite sums on  $k$ .

**Example 6.** Suppose we start with  $j = \$1000$ , bet \$100 at a time on Red in American roulette, and stop after 10 losses. Then  $p = 9/19$  and the payoff for winning is also \$100. On average, it takes  $E[T_{10}] = 19$  bets to stop. The final average fortune is  $E[\tilde{X}_{T_{10}}] = 1000 - 100 \times 10 + 100 \times 10 \times 9/10 = \$900$ ; thus,  $E[\tilde{X}_{(T_{10}-1)}] = \$1000$ . The average maximum fortune attained is  $E[\tilde{M}_{T_{10}}] = \$1287.55$ , the average minimum before the last loss is  $E[\tilde{m}_{(T_{10}-1)}] = \$712.45$ , while  $E[\tilde{m}_{T_{10}}] = \$690.25$ . Again we see that  $E[\tilde{M}_{T_n}] + E[\tilde{m}_{(T_n-1)}] - j = E[\tilde{X}_{(T_n-1)}]$  as in Corollary 3.

This roulette example reveals another interesting result. In this case,  $E[\tilde{M}_{T_{10}}]$  and  $E[\tilde{m}_{(T_{10}-1)}]$  are symmetric about the initial value being of the form  $1000 \pm 287.55$ . Indeed, for any  $n$ , this symmetry can happen with exactly one value of  $p$ . We state the result as follows:

**Theorem 6.** Let  $(\tilde{X}, j, p, c)$  be a random walk that moves up or down  $c$  units at a time. The average extreme values  $E[\tilde{M}_{T_n}]$  and  $E[\tilde{m}_{(T_n-1)}]$  are symmetric about the initial height of  $j$  if and only if  $p = (n - 1)/(2n - 1)$ .

*Proof.* We first note that the result in Theorem 3 can be adjusted to  $E[\tilde{M}_{T_n}] + E[\tilde{m}_{(T_n-1)}] = 2j - cn + cnp/q + c$ , while the average extrema are of the form  $E[\tilde{M}_{T_n}] = j + C_1$ ,  $E[\tilde{m}_{(T_n-1)}] = j - C_2$ . Then  $C_1 = C_2$  if and only if  $E[\tilde{M}_{T_n}] + E[\tilde{m}_{(T_n-1)}] = 2j$ , which means that  $-cn + cnp/q + c = 0$ , or  $-n + np/q + 1 = 0$ , so that that value  $c$  is irrelevant. Substituting  $q = 1 - p$  we have  $-n + np/(1 - p) + 1 = 0$ . Then solving for  $p$ , we obtain the single solution  $p = (n - 1)/(2n - 1)$ .  $\square$

With  $n = 10$ , then  $E[\tilde{M}_{T_{10}}]$  and  $E[\tilde{m}_{(T_{10}-1)}]$  are symmetric about the initial value  $j$  if and only if  $p = 9/19$  as when betting on Red in American roulette.

There are other average extrema that can be studied as well. For example, for  $m < j$ , let  $B_m$  be the number of steps needed to drop to height  $m$  from initial height  $j$ . Then the pdf of  $B_m$  and  $E[M_{B_m}]$  can be expressed in closed-form. Also for  $p + q < 1$ , we can allow for steps for which the height remains constant with probability  $r = 1 - p - q$ . Then  $T_n$  can be redefined to be the number of steps needed to obtain  $n$  downward or constant steps. The average area under the curve has been analyzed for this stopping time (see [4]). A much more challenging problem though is to find average extrema, when randomly stopping or after a fixed number of steps, when the upward and downward movements are not in a 1:1 ratio.

## References

- [1] L. Baringhaus, R. Grubel, Monte Carlo algorithms for finding the maximum of a random walk with negative drift, *Journal of Applied Probability*, **43** (2006), 74-86, doi: 10.1239/jap/1143936244.
- [2] W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume I, 3-rd Edition, John Wiley and Sons, USA (1968).
- [3] D. Korshunov, On distribution tail of the maximum of a random walk, *Stochastic Processes and their Applications*, **72**, No. 1 (1997), 97-103, doi: 10.1016/S0304-4149(97)00060-4.
- [4] M. Rogers, D. Neal, Generalized random walks with negative binomial stopping times, *New Zealand Journal of Mathematics*, **30** (2001), 69-79.
- [5] M. Russell, D. Neal, A generalized martingale betting strategy, *Missouri Journal of Mathematical Sciences*, **21**, No. 3 (2009), 183-197.

- [6] N. Veraverbeke, J.L. Teugels, The exponential rate of convergence of the distribution of the maximum of a random walk, *Journal of Applied Probability*, **12** (1975), 279-288, **doi:** 10.2307/3212441.

