Star Wars - an episode's battle

Patrícia Nunes da Silva (Corresponding author)
Mathematical Analysis, Rio de Janeiro State University,
Rua São Francisco Xavier, 524
Zip code 20550-900
Rio de Janeiro, RJ, Brazil.
nunes@ime.uerj.br
+ 55 21 98286 4346

Monica Almeida Gama
Escola Municipal Rui Barbosa,
Cabo Frio RJ, Brazil.

André Luiz Cordeiro dos Santos
The Federal Center for Technological Education Celso Suckow da Fonseca,
Rio de Janeiro, RJ, Brazil.

Abstract

Mlodinow (2008) proposed a crazy market experiment: to release the same film under two titles: Star Wars: Episode A and Star Wars: Episode B. Their marketing campaigns and distribution schedule are identical except by their titles on trailers and ads. He looks at the first 20,000 moviegoers and record the film they choose to see. He claims it is most probable the lead never changes, and it is 88 times more likely that one of the two films will be int the lead through all 20,000 customers than it is that the lead continuously seesaw. We present a detailed mathematical explanation for Mlodinow claims.

Keywords: misperceptions of randomness; random walk; discrete arc sin law; combinatorial methods;

1. Introduction

Mlodinow (2008) discusses many problems that defy our intuition and common sense. He is interested in mistaken judgments due to misperceptions of randomness or uncertainty. In one of his examples, Mlodinow (2008) proposed a crazy market experiment: to release the same film under two titles: Star Wars: Episode A and Star Wars: Episode B. Their marketing campaigns and distribution schedule are identical except by their titles on trailers and ads. He looks at the first 20,000 moviegoers and record the film they choose to see. To mathematically model his experiment, we are going to use random walks and paths. Combinatorial methods allow us to prove Mlodinow claims: it is most probable that the lead never changes, and 88 times more likely that one of the two films to be in the lead through all 20,000 customers than it is that each film to be in the lead among 10,000 moviegoers. Mlodinow experiment straightly relates to the classical fictitious gambler Peter, presented by Feller (1957). We combine Feller's (1957, 1968) and Border's (2017) results to prove Mlodinow's (2008) claim. In Border (2017), we have a comprehensive presentation of Feller's (1957, 1968) results. For completeness, we present their proofs.

2. Random walks and paths

A real Rademacher variable $X$is a random variable defined on some probability space which takes the values $+1$ or -1, each with probability $\frac{1}{2}$. That is $X: \Omega \to \{-1,1\}$ such that

$$P(X = +1) = P(X = -1) = \frac{1}{2}.$$

A Rademacher sequence is a sequence $(X_t)$of independent Rademacher random variables. The index t indicates an epoch. The set of epochs is the set E of non-negative integers. The epoch 0 is the moment before any vote.

For each t, we define the cumulative sum

$$S_t = X_1 + \cdots + X_t.$$

We define $S_0 = 0$. The sequence $S_0, S_1, S_2, \dots, S_t, \dots$ is a simple random walk on the integers.

Since the films and their marketing campaigns are the same, we can model the battle mathematically through a random walk. For the t-th viewer we associate a Rademacher variable X_t. If the viewer chooses Episode A, then $X_t = 1$; if the the viewer chooses Episode B, $X_t = -1$. To the sequence (X_t), we associate the random walk $S_{0,} S_{1,} S_{2,} \dots, S_t, \dots$.

2.1 Paths and reachable points

We associate the random walk (S_t) to a polygonal in the plane whose vertices are the points (t, S_t).

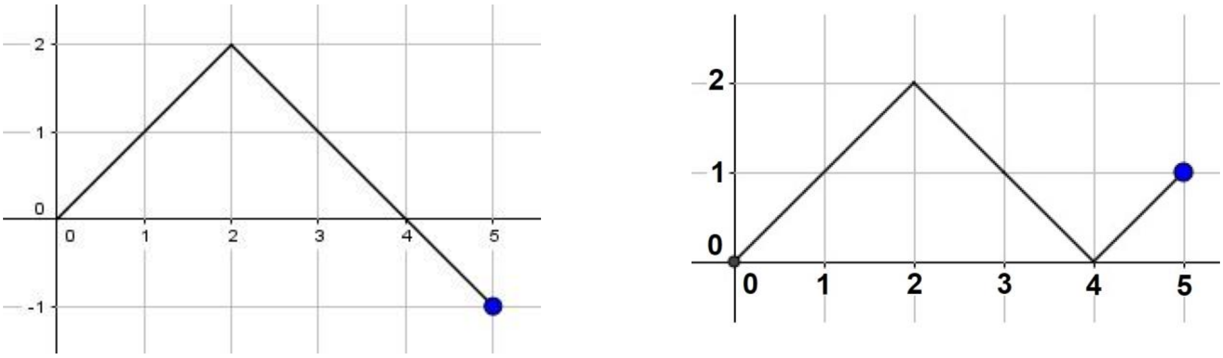
Definition 2.1. Let x, y be integers, $x > 0$. A path $s = (s_{0,} s_{1,} s_{2,} s_{3,} \dots, s_x)$ from the origin to the point (x, y) is the graph of the piecewise linear function whose vertices are (j, s_j), $j = 0,1,2, \dots, x$ satisfying:

$$s_j - s_{j-1} = \pm 1, \quad s_0 = 0, \quad s_x = y. \tag{1}$$

When we associate the random walk (S_t) with a path s, with $s_j = X_j$, we have a geometric representation of the battle between the episodes. In (1), we have

$$s_j - s_{j-1} = X_j = \pm 1, \qquad \text{for } j = 1,2, \dots, x$$

If the j-th vote was for episode A, we have $X_j = +1$; if it was for episode B, we have $X_j = -1$. The abscissa of the endpoint of a path represents the total number of votes (remembering that each person votes in one of the two episodes). If the partial sum $s_k = X_1 + X_2 + \cdots + X_k$ is positive, the partial result up to the k-th vote indicates the victory of episode A. When it is negative, it indicates the victory of episode B. In Figure 1, five viewers voted. In Figure 1(a), three viewers viewers chose episode A and two, B. In Figure 1(b), two viewers chose episode A and three, B.



(a) (b)

Figure 1. Five viewers vote.

Suppose that at the end of the vote, a viewers chose episode A and b, episode B, we have $x = a + b$ and $s_x = a - b$. If x is positive, A wins the contest. If the sum is negative, B wins. Any sequence of votes given to the episodes corresponds to a path and vice versa. Whenever the path is above the x-axis, it means that A has more cumulated votes than B. Similarly, in case B is winning, the polygonal will be below the x-axis. Thus, Episode A is in the lead as long as the polyline is above the x-axis. When it is below the x-axis, B is in the lead . A tie occurs when the polyline touches the x-axis. The cumulated sum of votes for each episode equals zero. That is, each episode received the same amount of votes. In this case, there has not yet been a change in the leading position. We assume that episode A is in the lead until time t if $S_t > 0$ or $S_t = 0$ and $S_{t-1} > 0$. Similarly, episode B is in the lead until time t if $S_t < 0$ or $S_t = 0$ and $S_{t-1} < 0$. That is, in the event of a tie, the episode that was ahead in the penultimate epoch is the lead.

When $s_t = k$, for $k \in \mathbb{Z}$, we say that the path s visits k at epoch t. If there is a path s such that $s_t = k$, we say that path s reaches the point (t, k) or that the point (t, k) is reachable from the origin.

The Proposition 2.1 tells us which points on the plane belonging to $E \times \mathbb{Z}$ are reachable.

Proposition 2.1 (Border (2017))**.** For the point (t, k) to be reachable, there must be non-negative integers a and b, such that

$$\begin{cases} a + b = t \\ a - b = k \end{cases} . \tag{2}$$

Note that not every point in the plan is reachable. For example, the point $(5,4)$ is not reachable. In fact, for $(t, k) = (5,4)$, the system has no solution (a, b) with a and b non-negative integers. That is, episode A cannot win by four votes if there are only five viewers.

Corollary 2.1. If $(t, k) \in E \times \mathbb{Z}$ is reachable, then the coordinates t and k have the same parity. Also,

$t \geq |k|$.

Proof. Let (t,k) be reachable. As $t \in E$, there are two possibilities for t:

1. $t$is even. That is, $t = 2v$ for some $v \in E$. We know by Proposition 2.1 that there is an non-negative integer solution of (2):

   $$2v = a + b \quad \text{and} \quad k = a - b$$

   Then $k = 2(a - v) = 2(v - b)$. Therefore, if t is even, k must also be even so that (2) has a non-negative integer solution. In this case

   $$a = \frac{t+k}{2} \quad \text{e} \quad b = \frac{t-k}{2}.$$

2. t is odd. That is, $t = 2v + 1$ for some $v \in E$. Similarly, by Proposition 2.1, there is a non-negative integer solution (a,b) of (2). Then $k = 2(a - v) - 1 = 2(v - b) + 1$. Therefore, if t is odd, k must also be odd so that (2) has a non-negative integer solution. Again

   $$a = \frac{t+k}{2} \quad \text{e} \quad b = \frac{t-k}{2}.$$

Since $t = a + b, a, b \geq 0$ and $k = a - b$, we have $|k| = |a - b| \leq max\{a, b\} \leq t$.

Definition 2.2. $N_{t,k}$ denotes the number of paths from origin to a point (t,k). If (t,k) is not reachable, then $N_{t,k} = 0$.

In Proposition 2.2, we calculate the number of different votes cast by t spectators ending with $S_t = k$.

Proposition 2.2 (Feller (1968)). If (t,k) is an achievable point, then

$$\mathrm{N}_{t,k} = \binom{t}{\frac{t+k}{2}} = \binom{t}{\frac{t-k}{2}}. \tag{3}$$

Proof. Since (t,k) is a reachable point, by Proposition 2.1, there are non-negative integers a and b that satisfy (2). Episode A received a votes. That is, a is the number of times $+1$ has occurred; b is the number of times -1 appears. When voting, $+1$ and -1 can appear in any order. That is, we have a permutation with repetition of $t = a + b$. Thus:

$$\mathrm{N}_{t,k} = \frac{(a+b)!}{a!b!} = \binom{a+b}{a} = \binom{a+b}{b} = \binom{t}{\frac{t+k}{2}} = \binom{t}{\frac{t-k}{2}}.$$

So there are precisely $N_{t,k}$ different paths from the origin to the point (t,k) corresponding to the possible votes during the contest.

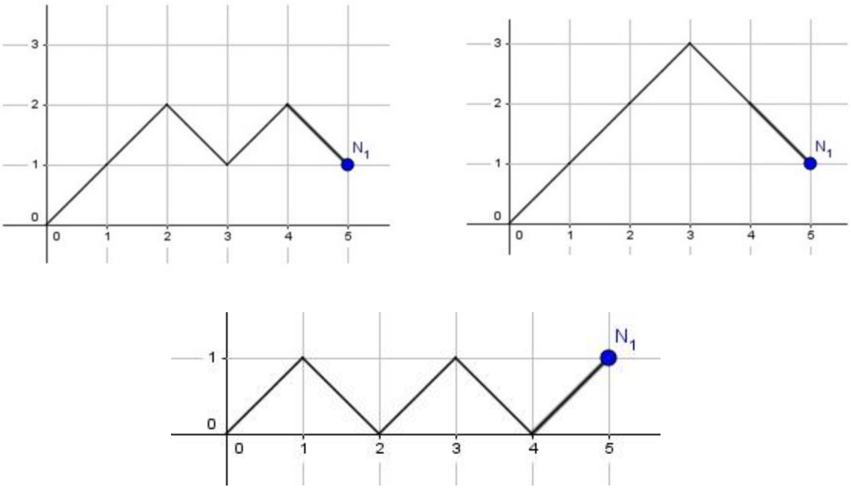


Figure 2. Episode A wins by one vote.

Figure 2 illustrates three scenarios of the episode A winning by one vote when five viewers voted. That is $(t,k)=(t,S_t)=(5,1)$. Ten different paths connect the origin to the point $(5,1)$.

On a path from origin to (t,k), We shall refer to t as the length of the path. For t, if we let k vary in (t,k), there are 2^t paths of length t. In fact, for each spectator's vote, there are two possible choices. By the multiplicative principle, after t votes, we obtain one of the 2^t possible paths. All paths are equally likely.

The event "at epoch t the vote difference between episodes is k" will be denoted by $\{S_t=k\}$. For its probability $P(S_t=k)$, we write $p_{t,k}$.

Corollary 2.2. If (t,k) is a reachable point, then

$$p_{t,k} = \binom{t}{\frac{t+k}{2}} \cdot 2^{-t}. \tag{4}$$

2.1.1 Special paths

We remember that episode A is in the lead until epoch t if $S_t>0$ or $S_t=0$ and $S_{t-1}>0$. Similarly, episode B is in the lead until epoch t if $S_t<0$ or $S_t=0$ and $S_{t-1}<0$. That is, in the event of a tie, the episode that was ahead in the penultimate epoch is in the lead.

- Let Z_t be the set of paths in which the two episodes received the same amount of votes. That is, the set of paths s where $s_t=0$.
- Let P_t be the set of paths in which episode A has always been in the lead, and there has never been a tie. That is, the set of paths s that satisfy $s_1>0,\dots,s_t>0$.
- Let N_t be the set of paths in which episode A has always been in the lead. That is, the set of paths s that satisfy $s_1\geq 0,\dots,s_t\geq 0$.

We now show some useful relationships between the sets Z_t, P_t and N_t.

Lemma 2.1 (Border (2017))**.** There is a one-to-one correspondence between P_{2m} and N_{2m-1}.
Proof. For all path s of P_{2m}, the first vote was for Episode A. In fact, as $s_1 > 0$, path s passes through $(1,1)$. In addition, we also have $s_j \geq 1$ for $j = 1, \dots, 2m$, because in P_{2m}, all partial sums are positive.

Let us consider point $(1,1)$ as the new origin of the Cartesian plane (Figure 3).

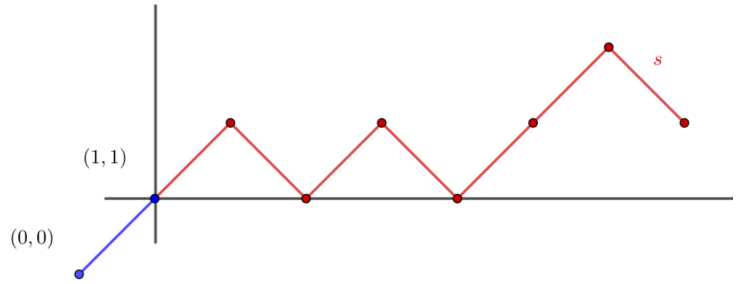


Figure 3. Path s.

In this way, we create a path s' of length $2m-1$ from s (Figure 4). The first vote was counted, $2m-1$ are left. Precisely, $s' = (s_0', s_1', \dots, s_{2m-1}') = (s_1 - 1, s_2 - 1, \dots, s_{2m} - 1)$. Thus, $s' \in N_{2m-1}$.

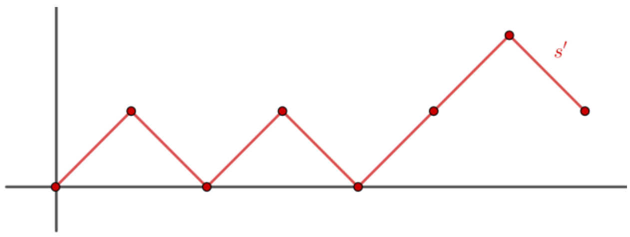


Figure 4. Path s'.

Similarly, with each s' in N_{2m-1}, we can always associate a single path $s \in P_{2m}$.
Lemma 2.2 (Nelson's lemma)**.** There is a one-to-one correspondence between Z_{2m} and N_{2m}. Furthermore, each path in Z_{2m} that has a minimum value of $-k$, corresponds to a path in N_{2m} that ends in $(2m, 2k)$.

Proof. (Border (2017)) To prove the lemma, we indicate how to build a bijective function $F: Z_{2m} \to N_{2m}$. Consider a path s in Z_{2m}. Since $s_{2m} = 0$, there is necessarily $j$such that $s_j \leq 0$. At some point $t \leq 2m$, it assumes a minimum value $-\hat{k} \leq 0$. Possibly $-\hat{k}$ is assumed more than once. Let $\hat{t}$ be the smallest t for which $s_t = -\hat{k}$.
Note that if the path s is already an element of N_{2m}, we have $s_t \geq 0$ for $t = 0, \dots, 2m$. Consequently, $\hat{k} = 0$ and $\hat{t} = 0$. In this case, we define $F(s) = s$. If s does not belong to N_{2m}, we necessarily have $s_t < 0$ for some $0 < t < 2m$. Then, $\hat{k} > 0$ and $0 < \hat{t} < 2m$ (Figure 5).

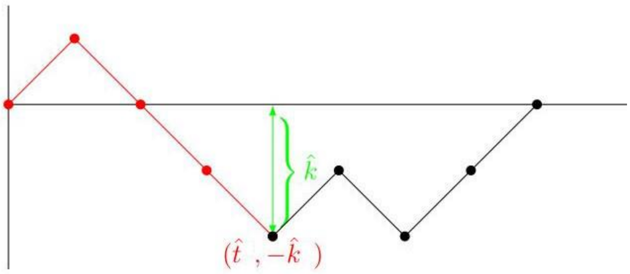


Figure 5. s does not belong to N_{2m}

To get a path s' in N_{2m}, we proceed as follows (see Figure 6):

- Take the section of path s from $(0,0)$ to $(\hat{t}, -\hat{k})$.
- Reflect this section over the vertical line $t = \hat{t}$.
- Slide the reflected section until the old endpoint $(\hat{t}, -\hat{k})$ coincides with the point $(2m, 0)$.
- Consider $(\hat{t}, -\hat{k})$ as the new origin of the Cartesian plane.

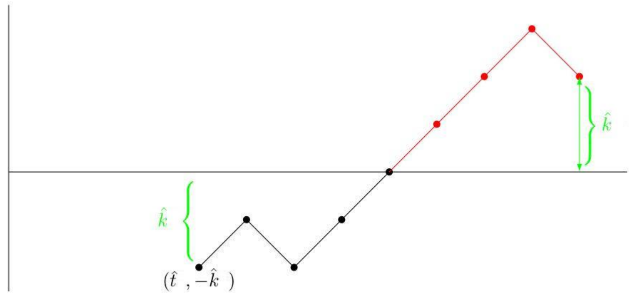


Figure 6. Path s' .

The path s' belongs to N_{2m}, and we define $s' = F(s)$ (see Figure 7).

$$s' = \left(s_{\hat{t}} + \hat{k}, s_{\hat{t}+1} + \hat{k}, \ldots, s_{2m} + \hat{k}, s_{\hat{t}-1} + 2\hat{k}, \ldots, s_1 + 2\hat{k}, 2\hat{k}\right).$$

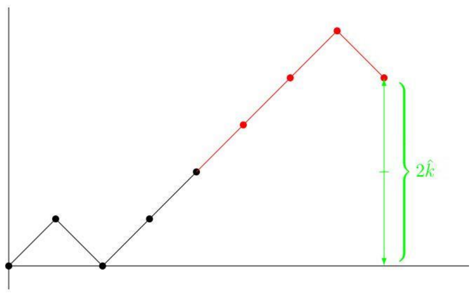


Figure 7. Path s' .

We now show that F is invertible: let s be a path in N_{2m}. If $s_{2m} = 0$, we have $F^{-1}(s) = s$. If $s_{2m} > 0$, we know from Corollary 2.1 that s_{2m} is even. That is, $s_{2m} = 2\bar{k}$, for some integer $\bar{k} > 0$. Consider $\bar{t}$ the last epoch when $s_t = \bar{k}$ (see Figure 8)

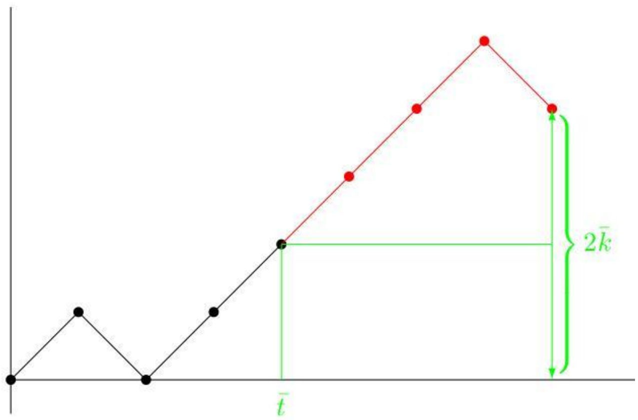
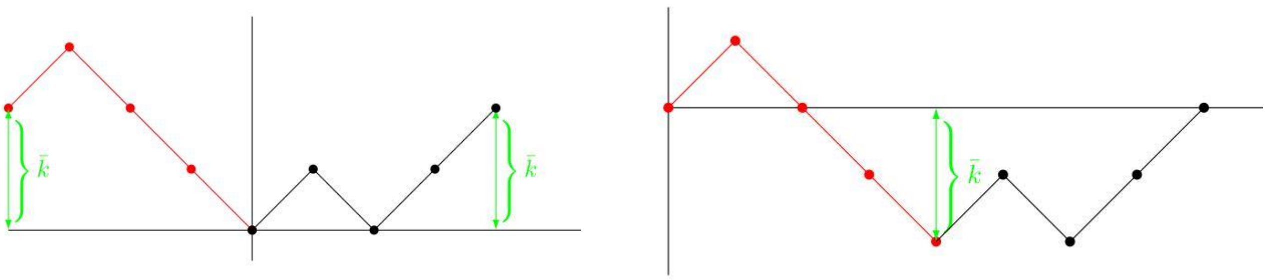


Figure 8. Last epoch when $s_t = \bar{k}$.

To get a path s' in Z_{2m}, we proceed as follows:

- Take the section of path s that runs from $(\bar{t}, \bar{k})$ to $(2m, 2\bar{k})$.
- Reflect this section over the vertical line $t = \bar{t}$.
- Slide the reflected section until the old endpoint $(\bar{t}, \bar{k})$ matches the origin (see Figure 9(a)).
- Consider the starting point as the new origin of the Cartesian plane (see Figure 9(b)).



(b)

(a)

Figure 9. The path $s' = F^{-1}(s)$.

The path s' belongs to Z_{2m}:

$$s' = \left(s_{2m} - 2\bar{k}, s_{2m-1} - 2\bar{k}, \dots, s_{\bar{t}} - 2\bar{k}, s_1 - \bar{k}, \dots, s_{\bar{t}-1} - \bar{k}, 0\right).$$

The procedure described for s' construction reverses that described in the definition of F. Thus thre is a one by one correspondence between Z_{2m} and N_{2m}.

2.1.2 Special events

Let's calculate the probability of some events of interest in the episode battle:

- $H = \{S_{2m} = 0\}$, when both episodes received the same amount of votes.

- $H = \{S_1 \geq 0, \ldots, S_{2m} \geq 0\}$, episode A has always been in the lead.
- $H = \{S_1 \leq 0, \ldots, S_{2m} \leq 0\}$, episode B has always been in the lead.

Lemma 2.3. The odds are identical:

$$P(S_{2m} = 0), \tag{5}$$

$$P(S_1 \geq 0, \ldots, S_{2m} \geq 0), \tag{6}$$

$$P(S_1 \leq 0, \ldots, S_{2m} \leq 0). \tag{7}$$

Proof. (Border (2017)) Note that all the events mentioned can be associated with paths of length $2m$. To calculate the probability $P(H)$ of an event H related to paths of length $2m$, it is necessary to determine the number of paths associated with event H and divide by 2^{2m}.

The probability in (6) is associated with the situation in which Episode A always wins. Draws can occur, but A is still in the lead, as there is no $S_t < 0$. The associated path does not cross the x-axis. Thus, there is no change in lead. In probability (7), it is the same situation, but the one who always wins is B. In both cases, by symmetry[1], the sets have the same cardinality. That is

$$|\{s, s_1 \geq 0, \ldots, s_{2m} \geq 0\}| = |\{s, s_1 \leq 0, \ldots, s_{2m} \leq 0\}|.$$

Therefore, $P(S_1 \geq 0, \ldots, S_{2m} \geq 0) = P(S_1 \leq 0, \ldots, S_{2m} \leq 0)$.
To conclude the lemma's proof, we show that the probabilities in (5) and (6) are equal. By Nelson's lemma (Lemma 2.2), we have $|Z_{2m}| = |N_{2m}|$. So $P(S_{2m} = 0) = P(S_1 \geq 0, \ldots, S_{2m} \geq 0)$.

3. Draws

We want to know if there was a change in the lead in the battle between the episodes. Thus, it is necessary to count the number of times that the associated paths crossed the x-axis. First, it is required to count how many times they have touched the mentioned axis. We now study the paths that connect the origin to an N point on the x-axis.

Definition 3.1 (Returns to zero)**.** When a path touches the x-axis, we say it returns to zero or the origin. In this case, $s_t = 0$ for some time t.

To return to the origin, the episodes must receive the same amount of votes. It follows from Corollary 2.1 that t is even. Consider $t = 2n$. The number of paths from the origin to $(2n, 0)$ is $N_{2n,0}$, so by Corollary 2.2, the probability u_{2n} of a path of length $2n$ ending at the point $(2n, 0)$ is given by:

$$u_{2n} = \binom{2n}{n} \cdot 2^{-2n}, \tag{8}$$

u_{2n} is the probability of a tie at epoch $t = 2n$.

[1] Just define $F: \{s, s_1 \geq 0, \ldots, s_{2m} \geq 0\} \rightarrow \{s, s_1 \leq 0, \ldots, s_{2m} \leq 0\}$, $F(s) = -s$.

Definition 3.2 (First return to zero)**.** The first return to zero occurs when a path touches the x-axis at the epoch $2m$ and$s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-1} \neq 0$. We denote by f_{2m} the probability of occurring the first return to zero in the epoch $2m$. That is,

$$f_{2m} = P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m-1} \neq 0, S_{2m} = 0).$$

Lemma 3.1 (Border (2017))**.** An explicit formula for f_{2m} is

$$f_{2m} = u_{2m-2} - u_{2m} = \frac{1}{2m-1} u_{2m} = \frac{1}{2m-1} \binom{2m}{m} \frac{1}{2^{2m}}. \tag{9}$$

Proof. As $2m-1$ is odd, by Corollary 2.1, we necessarily have $s_{2m-1} \neq 0$. So we can write the first return event at the epoch $2m$ as $\{S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m-2} \neq 0, S_{2m} = 0\}$. The paths associated with this event belong to the difference between two sets: of events in which s_t never vanishes until epoch $2m-2$ minus the events in which s_t never vanishes until epoch $2m$. That is,

$$\{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-2} \neq 0, s_{2m} = 0\} = \{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-2} \neq 0\} \setminus \{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m} \neq 0\}.$$

Note that for paths of length $2m$, we have

$$\{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m} \neq 0\} \subset \{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-2} \neq 0\}.$$

Besides, the number of paths of length $2m$ such that $s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-2} \neq 0$ is equal to four times the amount of paths of length $2m-2$ such that $s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-2} \neq 0$.

By Lemma 2.3, for $\varDelta = P\big((S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m-2} \neq 0) \setminus (S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m} \neq 0)\big)$, we have

$$\varDelta = \frac{4 \cdot |\{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m-2} \neq 0\}| - |\{s, s_1 \neq 0, s_2 \neq 0, \ldots, s_{2m} \neq 0\}|}{2^{2m}} = u_{2m-2} - u_{2m}$$

By (8), we have

$$u_{2m-2} - u_{2m} = \binom{2m-2}{m-1} \cdot 2^{-2(m-1)} - 2^{-2m} = \frac{1}{2m-1} \binom{2m}{m} \frac{1}{2^{2m}}.$$

We can also obtain a recursive formula for the return to zero involving the first returns.

Corollary 3.1 (Feller (1968))**.** For $m \geq 1$, we have

$$u_{2m} = \sum f_{2r}\, u_{2m-2r}. \tag{10}$$

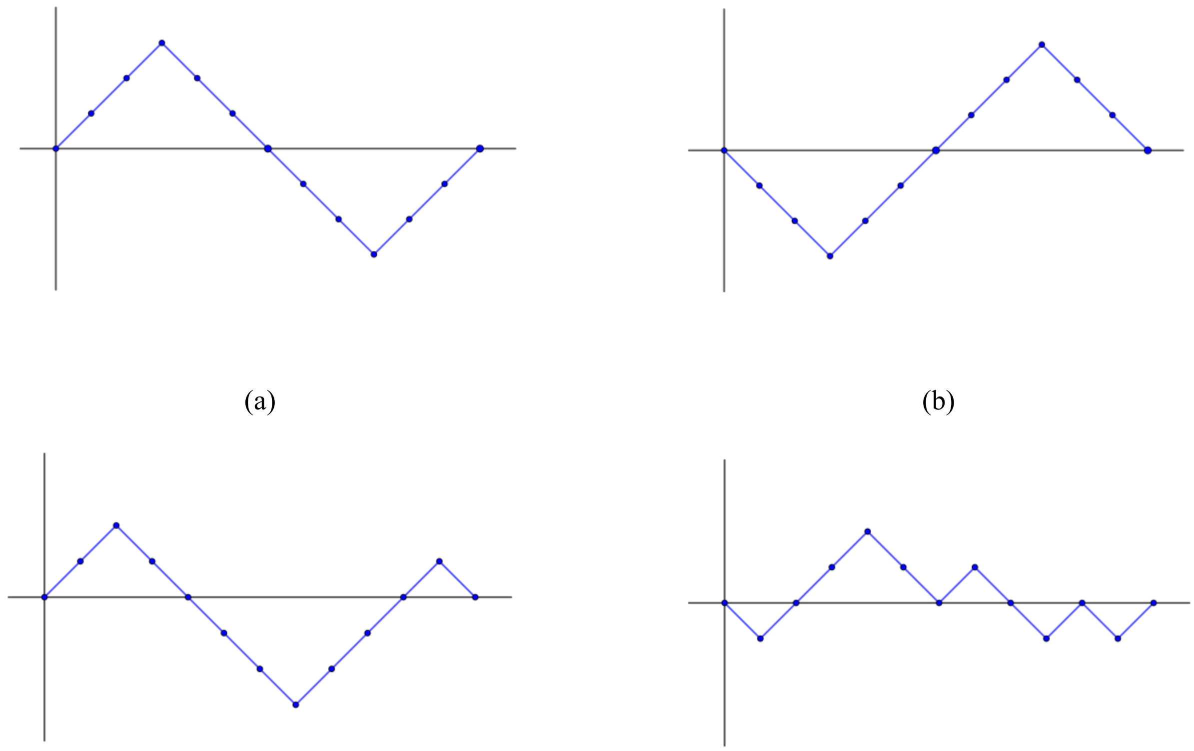
Proof. If a return to the origin occurs in the epoch $2m$, then the first return to zero occurs in an epoch $2r \leq 2m$. Every such path s has a section of length $2m$ where $s_1 \neq 0, s_2 \neq 0, \ldots, s_{2r-1} \neq 0$ and a section of length $2m-2r$ where$s_1 \neq 0, s_2 \neq 0, \ldots, s_{2r-1} \neq 0$. Therefore, the number of paths of length $2m$ from the origin to the point $(2m, 0)$, whose first return to zero happened at the point $(2r, 0)$ is given by

$$2^{2r} \cdot f_{2r} \cdot 2^{2m-2r} \cdot u_{2m-2r}.$$

Adding over r, we get (10).

4. In the lead

It is important to note that the analysis made by Feller (1957) and adapted by Mlodinow (2008) for the battle of episodes is not interested in uninterrupted leads, but in studying how long an episode is in the lead. In the Figures, each episode remained half the time in the lead.



(a) (b) (c) (d)

Figure 10. Half the time in the lead

Our intuition leads us to believe that in the contest between the episodes, each of them would stay about half the time in the lead, and frequent changes do not occur. However, as the following result shows, our intuition played a trick with us. The next theorem analyzes the fraction of the total time that a path remains above the x-axis. That is, the probability of episode A to stay in the lead over a fraction of the total votes. Fractions close to 1 are the least likely.

Theorem 4.1. Let $\alpha_{2k,2n}$ be the probability that during the interval from 0 to $2n$, episode A remains for 2k votes in the lead, and episode B remains for $2n-2k$ votes in the lead. So

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}, \qquad k = 0,1,\ldots,n. \tag{11}$$

Proof (Feller (1957)). We prove by induction on n. First, let's deduce a recursive expression for $\alpha_{2k,2n}$, with $1 \leq k \leq n-1$, which also involves probabilities of first return to zero. If episode A remains for $2k$

votes in the lead and episode B remains for $2n-2k$ votes in the lead, there is a tie at some point. Let $2r$ be the time when the first return to zero occurred. As $1 \le k \le n-1$, we have $2r < 2n$. That is, the first tie must take place before the voting ends. Otherwise, only one episode would remain in the lead at all times. Thus, as the first tie takes place in epoch $2r$, the path s associated with the contest of the episodes belongs to one of two classes:

- In the first class, episode A led the entire range from 0 to $2r$.
- In the second class, episode B led the entire range from 0 to $2r$.

In the first class, we necessarily have $r \le k \le n-1$, and in the section of the path s after $(2r, 0)$, episode A remains in the lead by exactly $2k-2r$ more votes. Let's calculate how many paths of length $2n$ there are in the first class. Note that we have $2^{2r} \cdot f_{2r}$ paths of length $2r$, whose first return occurs at the time $2r$. By symmetry, in half of them, episode A leads all the time. Also, there are $2^{2n-2r} \cdot \alpha_{2k-2r,2n-2r}$ paths of length $2n-2r$ in which episode A remains for $2k-2r$ votes in the lead. Consequently, the total number of paths of length $2n$ in the first class is given

$$\frac{1}{2} \cdot 2^{2r} \cdot f_{2r} \cdot 2^{2n-2r} \cdot \alpha_{2k-2r,2n-2r} = \frac{1}{2} \cdot 2^{2n} \cdot f_{2r} \cdot \alpha_{2k-2r,2n-2r}.$$

On each of these paths, episode A remains for $2k$ votes in the lead. Therefore, episode B remains for $2n-2k$ votes in the lead.

In the second class, episode B led until season $2r$. By an analogous argument, we have $k \le n-r$, and the total number of paths of length $2n$ in the second class is given by

$$\frac{1}{2} \cdot 2^{2r} \cdot f_{2r} 2^{2n-2r} \cdot \alpha_{2k,2n-2r} = \frac{1}{2} \cdot 2^{2n} \cdot f_{2r} \cdot \alpha_{2k,2n-2r}.$$

On each of these paths, episode A remains for $2k$ votes in the lead. Therefore, episode B remains for $2n-2k$ votes in the lead.

Note that the classes are disjoint with each other and are also disjoint with the classes corresponding to different values of r. Thus,

$$\alpha_{2k,2n} = \tfrac{1}{2} \sum f_{2r} \cdot \alpha_{2k-2r,2n-2r} + \tfrac{1}{2} \sum f_{2r} \cdot \alpha_{2k,2n-2r}. \quad (12)$$

Let us now prove (11), by induction on n. For $n=1$, we only have two possible values for k: $k=0$ or $k=1$. According to Lemma 2.3, the odds are the same for episode A remain in the lead throughout the interval from 0 to $2n$ and for that both episodes receive the same amount of votes until the $2n$ season. So, using (8), we get

$$P(S_1 \geq 0, \dots, S_{2m} \geq 0) = P(S_{2m} = 0) = u_{2n} = \binom{2n}{n} \cdot 2^{-2n}.$$

Therefore, $\alpha_{2n,2n} = u_{2n}u_0 = u_{2n}$. Similarly, using (7) from Lemma 2.3, we obtain $\alpha_{0,2n} = u_0 u_{2n} = u_{2n}$. That is, each episode is just as likely to remain in the lead for the entire 0 to $2n$ interval. Thus, (11) is verified for $n = 1$.

Our induction hypothesis is

$$\alpha_{2k,2v} = u_{2k}u_{2v-2k}, \qquad v = 1,2,\dots,n-1, \qquad k = 0,1,\dots,n-1.$$

Note that in the first summation in (12), we have $n-k \leq n-r \leq n-1$. In the second summation, $k \leq n-r \leq n-1$. Using the induction hypothesis in (12), we obtain

$$\alpha_{2k,2n} = \frac{1}{2}\sum f_{2r} \cdot \alpha_{2k-2r,2n-2r} + \frac{1}{2}\sum f_{2r} \cdot \alpha_{2k,2n-2r}$$

$$\alpha_{2k,2n} = \frac{1}{2}\sum f_{2r} \cdot u_{2k-2r}u_{2n-2k} + \frac{1}{2}\sum f_{2r} \cdot u_{2k}u_{2n-2r-2k}$$

$$\alpha_{2k,2n} = \frac{1}{2} \cdot u_{2n-2k} \sum f_{2r} \cdot u_{2k-2r} + \frac{1}{2} \cdot u_{2k} \sum f_{2r} \cdot u_{2n-2r-2k}.$$

By (10),

$$\alpha_{2k,2n} = \frac{1}{2} \cdot u_{2n-2k}u_{2k} + \frac{1}{2} \cdot u_{2k}u_{2n-2k} = u_{2k}u_{2n-2k}.$$

We use Theorem 4.1 to determine the probability of episode A to be in the lead over a fraction of the total votes. $2k$ is the number of votes over which episode A remains in the lead. We remember that

$$u_{2n} = \binom{2n}{n} \cdot 2^{-2n}.$$

Let $2n = 20.000$ and $k = 0$, that is, episode A will never take the lead. Substituting in formula (11), we have

$$\alpha_{0;20,000} = \frac{20{,}000!}{10{,}000! \cdot 10{,}000!} \cdot \frac{1}{2^{20{,}000}} = \frac{20{,}000!}{(10{,}000!)^2 \cdot 2^{20{,}000}}.$$

For $2k = 10{,}000$, both films remain in the lead for the same period,

$$\alpha_{10{,}000;10{,}000} = \frac{(10{,}000!)^2}{(5{,}000!)^4 \cdot 2^{20{,}000}}$$

To verify[2] that one probability is about 88 times greater than the other:

$$\frac{20{,}000!}{(10{,}000!)^2 \cdot 2^{20{,}000}} \div \frac{(10{,}000!)^2}{(5{,}000!)^4 \cdot 2^{20{,}000}} = 20{,}000! \cdot \left(\frac{5{,}000!}{10{,}000!}\right)^4 \simeq 8{,}8625 \cdot 10 = 88{,}625.$$

The probability that one of the films will remain in the lead for the entire period of the choice of 20,000 viewers is almost 88 times greater than that of having equal lead times for both films.

2 We calculate factorials by https://www.calculatorsoup.com/calculators/discretemathematics/factorials.php.

7. References

[1] K. C. Border. "Simple Random Walk". 2017. Available at: http://www.math.caltech.edu/~2016-17/2term/ma003/Notes/Lecture16.pdf.

[2] W. Feller. An Introduction to Probability Theory and its Applications. John Wiley & Sons Inc., 1957.

[3] ________. An Introduction to Probability Theory and its Applications. John Wiley & Sons Inc., 1968.

[4] L. Mlodinow. The Drunkard's Walk. Pantheon Books, 2008.