What is the Arcsine Law?

Ethan Ackelsberg

July 19, 2018

"We shall encounter theoretical conclusions which not only are unexpected but actually come as a shock to intuition and common sense. They will reveal that commonly accepted notions concerning chance fluctuations are without foundation and that the implications of the law of large numbers are widely misconstrued. For example, in various applications it is assumed that observations on an individual coin-tossing game during a long time interval will yield the same statistical characteristics as the observation of the results of a huge number of independent games at one given instant. This is not so...

"Contrary to popular opinion, it is quite likely that in a long coin-tossing game one of the players remains practically the whole time on the winning side, the other on the losing side."

–William Feller, Chapter III in [Fel68]

1 The Arcsine Law for Simple Random Walk

We will follow the treatment in [Dunb] for the proof of the arcsine law. Let $(\xi_i)_{i\in\mathbb{N}}$ be i.i.d. random variables with $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = \frac{1}{2}$, and let

$$S_n := \sum_{i=1}^{n} \xi_i.$$

For $n \in \mathbb{N}$, let

$$\begin{aligned} L_n &:= |\{1 \leq k \leq n : S_k > 0\}| \\ L'_n &:= |\{1 \leq k \leq n : S_k > 0 \text{ or } S_{k-1} > 0\}|. \end{aligned}$$

We want to know the distribution of L_n as $n \to \infty$. First we will deal with L'_n for combinatorial calculations. Note that $L'_{2n} = 2\,|\{1 \leq k \leq n : S_{2k-1} > 0\}|$.

Let $p_{n,k} = \mathbb{P}(L'_{2n} = 2k)$, and let $u_k = \mathbb{P}(S_{2k} = 0) = \binom{2k}{k}2^{-2k}$.

Proposition 1.

$$p_{n,k} = u_k u_{n-k} = 2^{-2n}\binom{2k}{k}\binom{2(n-k)}{n-k}$$

.

We will come back to the proof of this proposition soon. First, let's see what it gives us. Recall Stirling's formula:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Then we have

$$\begin{aligned}
u_k &= 2^{-2k}\frac{(2k)!}{(k!)^2} \\
&\sim 2^{-2k}\frac{\sqrt{4\pi k}(2k/e)^{2k}}{2\pi k(k/e)^{2k}} \\
&= \frac{1}{\sqrt{\pi k}}.
\end{aligned}$$

Corollary 1.1. *For $0 \le x \le 1$,*

$$\mathbb{P}\left(\frac{L'_{2n}}{2n} < x\right) \to \frac{2}{\pi}\arcsin\sqrt{x}$$

Proof.

$$\begin{aligned}
\mathbb{P}\left(\frac{L'_{2n}}{2n} < x\right) &= \sum_{k<xn} p_{n,k} \\
&\approx \frac{1}{\pi}\sum_{k<xn}\frac{1}{\sqrt{k(n-k)}} \\
&= \frac{1}{\pi}\frac{1}{n}\sum_{k/n<x}\frac{1}{\sqrt{(k/n)(1-k/n)}} \\
&\to \frac{1}{\pi}\int_0^x \frac{1}{\sqrt{x(1-x)}\ dx} \\
&= \frac{2}{\pi}\arcsin\sqrt{x}.
\end{aligned}$$

□

Definition 2. A random variable X is *arcsine distributed* if for $x \in [0,1]$,

$$\mathbb{P}(X \le x) = \frac{2}{\pi}\arcsin\sqrt{x}.$$

The density function on $[0,1]$ is given by

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

1.1 Proof of Proposition

In order to prove the proposition, we will use a number of combinatorial results from [Duna]. For several combinatorial arguments in this section, we will use the *reflection principle* for counting paths. The reflection principle says that for $a, b > 0$ and $x_0 < x_1$, the number of paths between (x_0, a) and (x_1, b) that touch the x-axis is the same as the total number of paths from $(x_0, -a)$ to (x_1, b). The bijection between these two sets of paths is given by reflecting the initial segment of a path (before the first intersection with the x-axis) over the x-axis (see Figure 2).

Theorem 3 (Positive Walks)**.**

$$\mathbb{P}(L_{2n} = 2n) = 2^{-(2n+1)}\binom{2n}{n}.$$

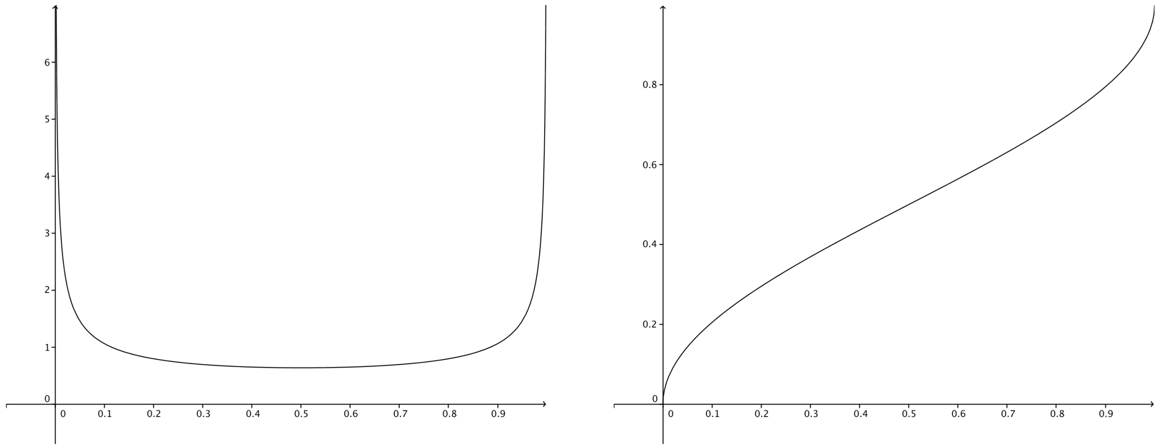

Figure 1: Graph of the density function (left) and the cumulative distribution function (right) for an arcsine distributed random variable.

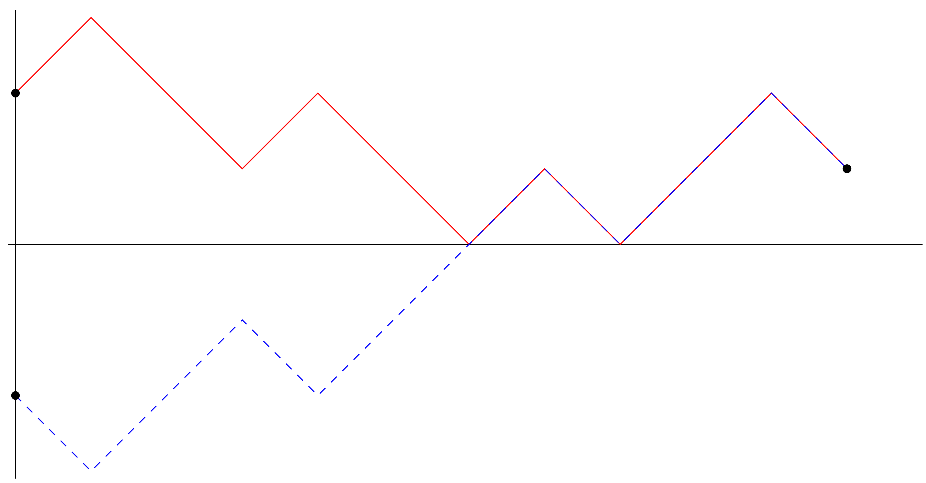

Figure 2: A path between $(0, 2)$ and $(11, 1)$ touching the x-axis (red path) and the corresponding reflected path from $(0, -2)$ to $(11, 1)$ (dotted blue path).

Proof. We want to count the paths from $(1, 1)$ to $(2n, 2k)$ that do not touch the x-axis. For $k = n$, there is exactly one such path. For $1 \leq k < n$, we will count the total number of paths between the endpoints minus those that touch the x-axis. The total number of paths from $(1, 1)$ to $(2n, 2k)$ is $\binom{2n-1}{n+k-1}$. By the reflection principle, the number of paths from $(1, 1)$ to $(2n, 2k)$ that touch the x-axis is the same as the number of paths from $(1, -1)$ to $(2n, 2k)$, which is $\binom{2n-1}{n+k}$. Thus, the number of paths beginning and the origin and staying in the upper half-plane is

$$1 + \sum_{k=1}^{n-1} \left(\binom{2n-1}{n+k-1} - \binom{2n-1}{n+k} \right) = 1 + \binom{2n-1}{n} - \binom{2n-1}{2n-1} = \binom{2n-1}{n}.$$

Finally,

$$\begin{aligned}\binom{2n-1}{n} &= \frac{(2n-1)!}{n!(n-1)!}\\ &= \frac{1}{2}\frac{2n(2n-1)!}{n!n(n-1)!}\\ &= \frac{1}{2}\binom{2n}{n}.\end{aligned}$$

□

Theorem 4 (Nonnegative Walks)**.**

$$\mathbb{P}(L'_{2n} = 2n) = 2^{-2n}\binom{2n}{n}.$$

Proof. We now want to count the number of paths of length $2n$ that are never negative. Such paths are in two-to-one correspondence with the paths from the previous theorem: shift the previous path over one and down one and then add a segment going either up or down. □

Theorem 5.

$$\mathbb{P}(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 0) = \frac{1}{n2^{2n}}\binom{2n-2}{n-1}.$$

Proof. We want to count the paths from $(1,1)$ to $(2n-1,1)$ that never touch the x-axis. This is the same as the number of nonnegative walks of length $2(n-1)$ beginning and ending at zero (Dyck path), so it is given by the Catalan number $C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1}$. (We can again use the reflection principle for this: take the total number of paths from $(1,1)$ to $(2n-1,1)$ minus those that touch that x-axis, which is the same as paths from $(1,-1)$ to $(2n-1,1)$.) □

We can now prove Proposition 1.

Proof of Proposition 1. The case $k = n$ is the nonnegative walks theorem.

We will induct on n for the general case $k < n$. Base case ($n = 1$):

$$\mathbb{P}(L'_{2n} = 0) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}.$$

Fix $N > 1$ and suppose the claim holds for $n \leq N-1$ and $k \leq n$. For $k = 0$, we have

$$\mathbb{P}(L'_{2N} = 0) = \mathbb{P}(\text{nonpositive walk}) = 2^{-2N}\binom{2N}{N}.$$

If $0 < L'_{2N} < 2N$, then there is a j for which $S_{2j} = 0$. Let $M := \inf\{j > 0 : S_{2j} = 0\}$. Then for $1 \leq k \leq N-1$, we have

$$\mathbb{P}(L'_{2N} = 2k) = \sum_{j=1}^{N} \mathbb{P}(L'_{2N} = 2k, M = j, S_1 > 0) + \sum_{j=1}^{N} \mathbb{P}(L'_{2N} = 2k, M = j, S_1 < 0).$$

Now, if $j > k$, then $\{L'_{2N} = 2k, M = j\} = \emptyset$. For $j \leq k$,

$$\begin{aligned}&\left|\{L'_{2N} = 2k, M = j, S_1 > 0\}\right|\\ &= |\{\text{paths from } (0,0) \text{ to } (2j,0) \text{ with } S_i > 0 \text{ for } i > 1\}|\\ &\times \left|\{\text{paths of length } 2(N-j) \text{ with } 2(k-j) \text{ segments in the upper half plane}\}\right|.\end{aligned}$$

The first term is $\frac{1}{j}\binom{2j-2}{j-1}$ by Theorem 5. The second term is $\binom{2(k-j)}{k-j}\binom{2(N-k)}{N-k}$ by induction. Hence,

$$\mathbb{P}(L'_{2n}=2k, M=j, S_1>0) = \frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2(k-j)}{k-j}\binom{2(N-k)}{N-k}.$$

Substituting back, we have

$$\begin{aligned}
\mathbb{P}(L'_{2N}=2k) &= \sum_{j=1}^{k}\frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2(k-j)}{k-j}\binom{2(N-k)}{N-k} \\
&+ \sum_{j=1}^{N-k}\frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2k}{k}\binom{2(N-k-j)}{N-k-j} \\
&= \left(2^{-2N}\binom{2(N-k)}{N-k}\right)\sum_{j=1}^{k}\frac{1}{j}\binom{2j-2}{j-1}\binom{2(k-j)}{k-j} \\
&+ \left(2^{-2N}\binom{2k}{k}\right)\sum_{j=1}^{N-k}\frac{1}{j}\binom{2j-2}{j-1}\binom{2(N-k-j)}{N-k-j} \\
&= \left(2^{-2N}\binom{2(N-k)}{N-k}\right)\frac{1}{2}\binom{2k}{k} + \left(2^{-2N}\binom{2k}{k}\right)\frac{1}{2}\binom{2(N-k)}{N-k} \\
&= 2^{-2N}\binom{2k}{k}\binom{2(N-k)}{N-k}.
\end{aligned}$$

Here we used the identity:

$$\sum_{j=1}^{n} C_{j-1}\binom{2(n-j)}{n-j} = \frac{1}{2}\binom{2n}{n}.$$

□

1.2 Arcsine Law

Theorem 6 (Arcsine Law/Law of Long Leads)**.**

$$\mathbb{P}\left(\frac{L_n}{n}<x\right) \to \frac{2}{\pi}\arcsin\sqrt{x}.$$

That is, $\frac{L_n}{n}$ *converges in distribution to the arcsine law.*

Proof. We will use the corollary and compare L_n to L'_n.

Note that $L_{2n} = L'_{2n} - |\{1 \le k \le n : S_{2k-1} > 0, S_{2k=0}\}|$, so

$$|L_{2n} - L'_{2n}| \le |\{k \le n : S_{2k}=0\}| = U_n.$$

We have

$$\mathbb{E}[U_n] = \sum_{k=1}^{n} u_k.$$

So by Markov's inequality,

$$\begin{aligned}\mathbb{P}(U_n > 2n\varepsilon) &\leq \frac{\mathbb{E}[U_n]}{n\varepsilon} \\ &= \frac{1}{2n\varepsilon}\sum_{k=1}^{n} 2^{-2k}\binom{2k}{k}.\end{aligned}$$

By Stirling, $2^{-2k}\binom{2k}{k}$ goes to zero at the rate of $k^{-1/2}$, so $\mathbb{P}(U_n > 2n\varepsilon) \to 0$.

Now,

$$\mathbb{P}(L_{2n} < 2nx) \leq \mathbb{P}(L'_{2n} < 2n(x+\varepsilon)) + \mathbb{P}(U_n > 2n\varepsilon).$$

Therefore,

$$\limsup_{n\to\infty} \mathbb{P}(L_{2n} < 2nx) \leq \frac{2}{\pi}\arcsin\sqrt{x+\varepsilon}.$$

Taking $\varepsilon \to 0$,

$$\limsup_{n\to\infty} \mathbb{P}(L_{2n} < 2nx) \leq \frac{2}{\pi}\arcsin\sqrt{x}.$$

In general, $L_{2n} \leq L'_{2n}$, so

$$\liminf_{n\to\infty} \mathbb{P}(L_{2n} < 2nx) \geq \frac{2}{\pi}\arcsin\sqrt{x}.$$

Finally,

$$\begin{aligned}&\mathbb{P}(L_{2n+1} < (2n+1)x) \leq \mathbb{P}(L_{2n} < (2n+1)x), \text{ and} \\ &\mathbb{P}(L_{2n+2} < (2n+2)x) \leq \mathbb{P}(L_{2n+1} < (2n+2)x),\end{aligned}$$

so the probabilities at odd times are appropriately sandwiched. □

Similar results hold for random walks with other i.i.d. centered steps [EK47, And54] and for Brownian motion [L3́9], which is a continuous analog of a random walk. Moreover, arcsine laws govern the time of the maximum and the time of the last zero for random walks.

2 Consequences

In a long sequence of coin flips, the probability that heads leads 85% of the time is about 25.32% (see Figure 3). This means that in over half of long, evenly matched games, one team will lead for at least 85% of the time, a surprisingly long amount.

Example from [Dunb]: An investment has a positive net fortune 75% of the time. This, however, has a 33% chance of happening if the investment is totally random with mean zero, so it may not be a worthwhile investment.

The paper [CKR15] shows that in NBA basketball games, statistics on leads follow the arcsine law, as would be expected if games behave like random walks. In particular, lead changes are most likely at the beginning and end of games and less likely in the middle, so "comebacks" are not so surprising.

If all students are equally capable but have some randomness in performance, it would not be surprising for one student to maintain the highest overall grade and another the lowest overall grade for the vast majority of the semester.

x	$\frac{2}{\pi}\arcsin\sqrt{x}$
0	0
0.05	0.1436
0.1	0.2048
0.15	0.2532
0.2	0.2952
0.25	0.3333
0.3	0.3690
0.35	0.4030
0.4	0.4359
0.45	0.4681
0.5	0.5

Figure 3: Values of $\frac{2}{\pi}\arcsin\sqrt{x}$.

3 Erdős Arcsine Law

The arcsine law also appears for statistical properties of prime factors. For $n \in \mathbb{N}$, write $n = p_1^{r_1} \cdots p_k^{r_k}$ with $p_1 < p_2 < \cdots < p_k$ prime. We say a factor p_j is *small* if $\log\log p_j < j$. Let $s_n = \#\{1 \le j \le k : p_j \text{ is a small prime factor of } n\}$.

Theorem 7 (Erdős, [Erd69])**.** *For $x \ge 0$,*

$$\frac{\{1 \le n \le N : s_n < x \log\log n\}}{N} \to \frac{2}{\pi}\arcsin\sqrt{x}.$$

That is, $\frac{s_n}{\log\log n}$ behaves like an arcsine distributed random variable.

References

[And54] E. Sparre Andersen. On the fluctuations of sums of random variables. *Mathematica Scandinavica*, 1954.

[CKR15] A. Clauset, M. Kogan, and S. Redner. Safe leads and lead changes in competitive team sports. *Physical Review E*, 2015.

[Duna] S. R. Dunbar. Topics in probability theory and stochastic processes: Positive walks.

[Dunb] S. R. Dunbar. Topics in probability theory and stochastic processes: The arcsine law.

[EK47] P. Erdős and M. Kac. On the number of positive sums of independent random variables. *Bulletin of the American Mathematical Society*, 1947.

[Erd69] P. Erdős. On the distribution of prime divisors. *Aequationes Mathematicae*, 1969.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 1968.

[Kac69] M. Kac. *Statistical Independence in Probability, Analysis, and Number Theory*. Mathematical Association of America, 1969.

[L3́9] P. Lévy. Sur certains processus stochastiques homogènes. *Compositio Mathematica*, 1939.