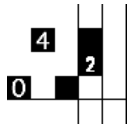


Steven R. Dunbar
Department of Mathematics
203 Avery Hall
University of Nebraska-Lincoln
Lincoln, NE 68588-0130
<http://www.math.unl.edu>
Voice: 402-472-3731
Fax: 402-472-8466

Topics in Probability Theory and Stochastic Processes Steven R. Dunbar

The Arcsine Law



Rating

Mathematicians Only: prolonged scenes of intense rigor.



Section Starter Question

In a coin-flipping game, do you expect the lead to change often? Graphically, how would you recognize a change in lead? What does the Weak Law of Large Numbers have to say about the lead changing often? What does the Central Limit Theorem have to say about the lead changing often?



Key Concepts

1. The **Arcsine Law**, (sometimes known as the **law of long leads**), says that in a coin-tossing games, a surprisingly large fraction of sample paths leave one player in the lead almost all the time, and in very few cases will the lead change sides and fluctuate in the manner that is naively expected of a well-behaved coin.
 2. Interpreted geometrically as random walks, the path crosses the x -axis rarely, and with increasing duration of the walk, the frequency of crossings decreases, and the lengths of the “leads” on one side of the axis increase in length.
-



Vocabulary

1. The **Arcsine Law**, (sometimes known as the **law of long leads**) says

$$\lim_{n \rightarrow \infty} \mathbb{P}_n [V_n < n\alpha] = \frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin \sqrt{\alpha}.$$



Mathematical Ideas

Heuristics for the Arcsine Law

Consider $T_n = Y_1 + \cdots + Y_n$ summing independent, identically distributed coin-toss random variables Y_i , each of which assumes the value $+1$ with probability $1/2$, and value -1 with probability $1/2$.

Recall that the stochastic process T_n is a function of two variables: the time n and the sample point ω . The Central Limit Theorem and the Moderate Deviations Theorem, give asymptotic results in n about the probability, that is, the proportion of ω values with an specific excess of heads over tails at that fixed n . That event could be expressed in terms of the event $\{\omega : T_n(\omega) > s(n)\}$ for some s . Now we are going to take a somewhat complementary point of view, asking about an event that counts the amount of time that the net fortune or walk is positive.

We say that the fortune spends the time from $k - 1$ to k on the positive side if at least one of the two values T_{k-1} and T_k is positive (in which case the other is positive or at worst, 0). Geometrically, the broken line path of the fortune lies above the horizontal axis over the interval $(k - 1, k)$.

For notation, let

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

Then u_{2n} is the binomial probability for exactly n heads and n tails in $2n$ flips of a fair coin.

Proposition 1. *Let $p_{2k,2n}$ be the probability that in the time interval from 0 to $2n$, the particle spends $2k$ time units on the positive side and therefore $2n - 2k$ on the negative side. Then*

$$p_{2k,2n} = u_{2k} u_{2n-2k}$$

We feel intuitively that the fraction k/n of the total time spent on the positive side is most likely to be $1/2$. However, the opposite is true! The middle values of k/n are least probable and the extreme values $k/n = 0$ and $k/n = 1$ are most probable in terms of frequency measured by the probability distribution!

The formula of the Proposition is exact, but not intuitively revealing. To make more sense, Stirling's Formula shows that

$$u_{2n} \sim \frac{\sqrt{2\pi(2n)} (2n)^{2n} e^{-2n}}{(\sqrt{2\pi n} n^n e^{-n})^2} \frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}}$$

as $n \rightarrow \infty$. Note that this application of Stirling's formula says the probability of n heads and n tails in $2n$ flips of a fair coin goes to 0 at the rate $1/\sqrt{n}$ as n gets large.

It then follows that

$$p_{2k,2n} \approx \frac{1}{\pi \sqrt{k(n-k)}}$$

as $k \rightarrow \infty$ and $(n-k) \rightarrow \infty$. The fraction of time that the fortune spends on the positive side is then k/n and the probability the fortune is on this positive side this fraction of the time is $p_{2k,2n}$. We can look at the cumulative probability that the fraction of time spent on the positive side is less than α (with $\alpha \leq 1$) namely,

$$\sum_{k < \alpha n} p_{2k,2n} \approx \frac{1}{\pi} \sum_{k < \alpha n} \frac{1}{\sqrt{(k/n)(1-k/n)}} \frac{1}{n}$$

On the right side we recognize a Riemann sum approximating the integral:

$$\frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin(\sqrt{\alpha}) dx.$$

For reasons of symmetry, the probability that $k/n \leq 1/2$ tends to $1/2$ as $n \rightarrow \infty$. Adding this to the integral, we get:

Theorem 2 (Arcsine Law). *For fixed α with $0 < \alpha < 1$ and $n \rightarrow \infty$, the probability that the fortune T_n spends a fraction of time k/n on the positive side is less than α tends to:*

$$\frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin(\sqrt{\alpha})$$

The Arcsine Law for Bernoulli Trials

Recall that Y_i is a sequence of independent random variables that take values 1 with probability $1/2$ and -1 with probability $1/2$. This is a mathematical model of a fair coin flip game where a 1 results from “heads” on the i th coin toss and a -1 results from “tails”. Define $T = (T_0, \dots, T_n)$ by setting $T_n = \sum_{i=0}^n Y_i$ with $T_0 = 0$.

A common interpretation of this probability game is to imagine it as a **random walk**. That is, we imagine an individual on a number line, starting at some position T_0 . The person takes a step to the right to $T_0 + 1$ with probability p and takes a step to the left to $T_0 - 1$ with probability q and continues this random process. Then instead of the total fortune at any time, we consider the geometric position on the line at any time.

Create a common graphical representation of the game. A continuous piecewise linear curve in \mathbb{R}^2 consisting of a finite union of segments of the form $[(i, j), (i + 1, j + 1)]$ or $[(i, j), (i + 1, j - 1)]$ where i, j are integers is called a **path**. A path has an origin (a, b) and an endpoint (c, d) that are points on the curve with integer coordinates satisfying $a \leq i \leq c$ for all (i, j) on the curve. The length of the path is $c - a$. (Note that the Euclidean length of the path is $(c - a)\sqrt{2}$.) To each element $\omega \in \Omega_n$ (see Binomial Distribution), we associate a path $\bigcup_{k=0}^{n-1} [(i, T_i(\omega)), (i + 1, T_{i+1}(\omega))]$ with origin $(0, 0)$ and endpoint $(n, T_n(\omega))$.

Definition. Set

$$V_n = |\{k : 0 \leq k \leq n, T_k > 0\}|.$$

This is the number of tosses in which the Heads player is ahead or winning. Let $V'_n = |\{k : 1 \leq k \leq n, T_k > 0 \text{ or } T_{k-1} > 0\}|$. This is the number of integers, or steps, between 1 and n inclusive such that there were more Heads than Tails in the first k or $k - 1$ tosses of the coin.

Line segments of a path are in the upper half plane if and only if $T_{2k-1} > 0$. Thus, we can say

$$V'_{2n} = 2|\{k : 1 \leq k \leq n \text{ and } T_{2k-1} > 0\}|.$$

Proposition 1. For each $n > 0$ and $0 \leq k \leq n$, then

$$\mathbb{P}_{2n} [V'_{2n} = 2k] = 2^{-2n} \binom{2k}{k} \binom{2(n-k)}{n-k}$$

Proof. Note that

$$\mathbb{P}_{2n} [V'_{2n} = 2n] = \mathbb{P}_{2n} [T_k \geq 0, k \in \{1, \dots, 2n\}] = 2^{-2n} \binom{2n}{n}.$$

by the Nonnegative Walks Theorem (Theorem 3 in Positive Walks Theorem). Prove the general statement of the current Proposition by induction. The base case where $n = 1$ is that

$$\begin{aligned} \mathbb{P}_2 [V'_2 = 0] &= 2^{-2(1)} \binom{2(0)}{0} \binom{2(1-0)}{1} \\ &= \frac{1}{2} \end{aligned}$$

which is clearly true since $\mathbb{P}_2 [V'_2] = \mathbb{P}_2 [T_2 > 0 \text{ or } T_1 > 0]$.

Fix $N > 1$, our inductive hypothesis is that the proposition is true for all $n \leq N - 1$ and for all $0 \leq k \leq n$. Note that if $k = 0$, we have

$$\mathbb{P}_{2N} [V'_{2N} = 0] = \mathbb{P}_{2N} [T_k \leq 0, 1 \leq k \leq N] = 2^{-2N} \binom{2N}{N}$$

again by the Nonnegative Walks Theorem (Theorem 3 in Positive Walks Theorem). If $0 < V'_{2N} < 2N$, then there exists j with $1 \leq j \leq N$ so that $T_{2j} = 0$. For each $\omega \in \Omega_{2N}$ such that $0 < V'_{2N}(\omega) < 2N$, the first time back to 0 is given by

$$v(\omega) = \min\{j > 0 : T_{2j}(\omega) = 0\}.$$

Fix $k \in \{1, \dots, N - 1\}$. Then

$$\begin{aligned} \mathbb{P}_{2N} [V'_{2N} = 2k] &= \sum_{j=1}^N \mathbb{P}_{2N} [V'_{2N} = 2k, v(\omega) = 2j, T_1 > 0] \\ &\quad + \sum_{j=1}^N \mathbb{P}_{2N} [V'_{2N} = 2k, v(\omega) = 2j, T_1 < 0] \end{aligned}$$

If $j > k$, then note that

$$\{V'_{2N} = 2k, v(\omega) = 2j, T_1 > 0\} = \emptyset.$$

Note that

$$\begin{aligned}
& |\{V'_{2N} = 2k, v(\omega) = 2j, T_1 > 0\}| \\
&= (\text{number of paths } (0, 0) \text{ to } (2j, 0) \text{ with } T_i > 0 \text{ for } i > 1) \\
&\times (\text{number of paths starting at } (2j, 0) \text{ and length } 2(N - j) \\
&\text{with } 2(k - j) \text{ elementary segments in the upper half plane})
\end{aligned}$$

The first term in the product is given by Corollary 1 in Positive Walks Theorem and is $\frac{1}{j} \binom{2j-2}{j-1}$ and the second term is

$$2^{2(N-j)} \mathbb{P}_{2N} [V'_{2(N-j)} = 2(k-j)] = \binom{2(k-j)}{k-j} \binom{2(N-k)}{N-k}.$$

Thus, we see that

$$\mathbb{P}_{2N} [V'_{2N} = 2k, t(\omega) = 2j, T_1 > 0] = \frac{1}{j 2^{2N}} \binom{2j-2}{j-1} \binom{2(k-j)}{k-j} \binom{2(N-k)}{N-k}.$$

Now combining the results we have

$$\begin{aligned}
& \mathbb{P}_{2N} [V'_{2N} = 2k] \\
&= \sum_{j=1}^k \frac{1}{2j^{2N}} \binom{2j-2}{j-1} \binom{2(k-j)}{k-j} \binom{2(N-k)}{N-k} \\
&\quad + \sum_{j=1}^{N-k} \frac{1}{2j^{2N}} \binom{2j-2}{j-1} \binom{2k}{k} \binom{2(N-j-k)}{N-j-k} \\
&= \left[\frac{1}{2^{2N}} \binom{2(N-k)}{N-k} \right] \sum_{j=1}^k \frac{1}{j} \binom{2j-2}{j-1} \binom{2(k-j)}{k-j} \\
&\quad + \left[\frac{1}{2^{2N}} \binom{2k}{k} \right] \sum_{j=1}^k \frac{1}{j} \binom{2j-2}{j-1} \binom{2(N-j-k)}{N-j-k} \\
&= \left[\frac{1}{2^{2N}} \binom{2(N-k)}{N-k} \right] \left(\frac{1}{2} \right) \binom{2k}{k} \\
&\quad + \left[\frac{1}{2^{2N}} \binom{2k}{k} \right] \left(\frac{1}{2} \right) \binom{2(N-k)}{N-k} \\
&= \frac{1}{2^{2N}} \binom{2k}{k} \binom{2(N-k)}{N-k},
\end{aligned}$$

which means that induction holds and so we have proven the Proposition \square

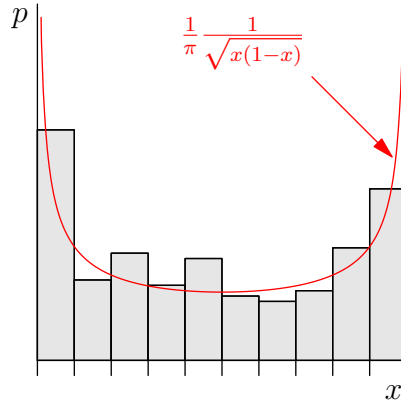


Figure 1: Illustration of the Arcsine Law using a simulation with 200 trials of random walks of length $n = 100$.

Theorem 3 (Arcsine Law). *For each $\alpha \in (0, 1)$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}_n [V_n < n\alpha] = \frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin \sqrt{\alpha}.$$

Proposition 4. *For all a, b with $0 \leq a \leq b \leq 1$, then*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{2n} [2na \leq V'_{2n} \leq 2nb] = \frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx.$$

Proof. 1. First, if we have $0 < a < b < 1$, then by Proposition 1 and Stirling's Approximation tell us

$$\begin{aligned} \mathbb{P}_{2n} [V'_{2n} = 2k] &= \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}} (1 + \epsilon(k)) (1 + \epsilon(n-k)) \\ &= \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}} (1 + \epsilon(n, k)). \end{aligned}$$

with $\lim_{n \rightarrow \infty} \epsilon(n, k) = 0$ uniformly in $k \in \mathbb{Z}$ for $na \leq k \leq nb$. Thus, we

have

$$\begin{aligned}
\mathbb{P}_{2n} [2na \leq V'_{2n} \leq 2nb] &= \sum_{na \leq k \leq nb} \mathbb{P}_{2n} [V'_{2n} = 2k] \\
&\sim \sum_{a \leq \frac{k}{n} \leq b} \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}} \\
&= \sum_{a \leq \frac{k}{n} \leq b} \frac{1}{\pi} \frac{1}{\sqrt{n^2 \left(\frac{k}{n}\right) \left(1 - \frac{k}{n}\right)}} \\
&= \sum_{k=0}^n \left(\chi_{[a,b]} \left(\frac{k}{n} \right) \right) \frac{1}{\pi n} \frac{1}{\sqrt{\left(\frac{k}{n}\right) \left(1 - \frac{k}{n}\right)}} \\
&= \frac{1}{n\pi} \sum_{k=0}^n \left(\chi_{[a,b]} \left(\frac{k}{n} \right) \right) \frac{1}{\sqrt{\left(\frac{k}{n}\right) \left(1 - \frac{k}{n}\right)}} \\
&\rightarrow \frac{1}{\pi} \int_0^1 \chi_{[a,b]} \frac{1}{\sqrt{x(1-x)}} dx \\
&= \frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx.
\end{aligned}$$

Note that here we actually have a Riemann sum since we have a bounded function when we keep $a \neq 0$ and $b \neq 1$. The rest of this proof is to allow $a = 0$ and $b = 1$.

2. Fix $\epsilon > 0$. There exists an a so that $\frac{1}{\pi} \int_0^a \frac{1}{\sqrt{x(1-x)}} dx < \epsilon$ and $\frac{1}{\pi} \int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}} dx < \epsilon$. From part 1 of the proof, we have

$$\left| \frac{1}{\pi} \int_a^{1-a} \frac{1}{\sqrt{x(1-x)}} dx - \mathbb{P}_{2n} [2na \leq V'_{2n} \leq 2n(1-a)] \right| < \epsilon$$

for n sufficiently large.

3. Note that $\mathbb{P}_{2n} [V'_{2n} < 2na] + \mathbb{P}_{2n} [2na \leq V'_{2n} \leq 2n(1-a)] + \mathbb{P}_{2n} [V'_{2n} > 2n(1-a)] = 1$. Note also that $\frac{1}{\pi} \int_0^1 \frac{1}{\sqrt{x(1-x)}} dx = 1$. Together these imply that

$$\frac{1}{\pi} \int_a^{1-a} \frac{1}{\sqrt{x(1-x)}} dx = 1 - \int_0^a \frac{1}{\sqrt{x(1-x)}} dx - \int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}} dx,$$

or in other words,

$$\mathbb{P}_{2n} [2na \leq V'_{2n} \leq 2n(1-a)] = 1 - \mathbb{P}_{2n} [V'_{2n} < 2na] - \mathbb{P}_{2n} [V'_{2n} > 2n(1-a)].$$

From these facts and part 2 we have

$$\begin{aligned} & \left| \left(\frac{1}{\pi} \int_0^a \frac{1}{\sqrt{x(1-x)}} dx \right. \right. \\ & \quad \left. \left. + \frac{1}{\pi} \int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}} dx - 1 \right) \right. \\ & \quad \left. + (1 - \mathbb{P}_{2n} [V'_{2n} < 2na] - \mathbb{P}_{2n} [V'_{2n} > 2n(1-a)]) \right| \end{aligned}$$

for sufficiently large n . Thus, for sufficiently large n have

$$|\mathbb{P}_P [2n] V'_{2n} < 2na + \mathbb{P}_{2n} [V'_{2n} > 2n(1-a)]| < 3\epsilon.$$

So there exists $a > 0$ so that $\mathbb{P}_{2n} [V'_{2n} < 2na] < 3\epsilon$ for sufficiently large n . Since $\mathbb{P}_{2n} [V'_{2n} < 2na]$ is increasing in a for fixed n , we get $\lim_{a \rightarrow \infty} \mathbb{P}_{2n} [V'_{2n} < 2na] = 0$ uniformly in n .

4. By parts 1 and 2, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_n [V'_{2n} \leq 2nb] = \frac{1}{\pi} \int_0^b \frac{1}{\sqrt{x(1-x)}} dx$$

for $b \in (0, 1)$. By symmetry, the proposition holds. \square

Proof of the Arcsine Law. The proof uses the relationship between the random variables V_{2n} and V'_{2n} . Since $V_{2n} := V'_{2n} - |\{k : 1 \leq k \leq n, T_{2k-1} > 0, T_{2k} = 0\}|$, it follows that

$$|V_{2n} - V'_{2n}| \leq |\{k : 1 \leq k \leq n, T_{2k} = 0\}| = U_{2n}. \quad (1)$$

Note that

$$\begin{aligned} \mathbb{E}_{2n} [U_{2n}] &= \mathbb{E}_{2n} \left[\sum_{k=1}^n \chi_{[T_{2k}=0]} \right] \\ &= \sum_{k=1}^n \mathbb{P}_{2n} [T_{2k} = 0] \\ &= \sum_{k=1}^n 2^{-2k} \binom{2k}{k}. \end{aligned}$$

By Markov's Inequality,

$$\begin{aligned}\mathbb{P}_{2n} [U_{2n} > 2n\epsilon] &\leq \frac{\mathbb{E}_{2n} [U_{2n}]}{2n\epsilon} \\ &= \frac{1}{2n\epsilon} \sum_{k=1}^n 2^{-2k} \binom{2k}{k}.\end{aligned}$$

By Stirling's Approximation, we have $\lim_{k \rightarrow \infty} 2^{-2k} \binom{2k}{k} = 0$ at the same rate as $\frac{1}{\sqrt{k}}$. Cesàro's Principle gives

$$\lim_{n \rightarrow \infty} \mathbb{P}_{2n} [U_{2n} > 2n\epsilon] = 0. \quad (2)$$

Now note that

$$[V_{2n} < 2n\alpha] \subset [|V_{2n} - V'_{2n}| > 2n\epsilon] \cup [V'_{2n} \leq 2n(\alpha + \epsilon)].$$

Thus, we have

$$\mathbb{P}_{2n} [V_{2n} < 2n\alpha] \leq \mathbb{P}_{2n} [|V_{2n} - V'_{2n}| > 2n\epsilon] + \mathbb{P}_{2n} [V'_{2n} \leq 2n(\alpha + \epsilon)]. \quad (3)$$

For the first probability on the right

$$\mathbb{P}_{2n} [|V_{2n} - V'_{2n}| > 2n\epsilon] \leq \mathbb{P}_{2n} [U_{2n} > 2n\epsilon] \rightarrow 0.$$

as $n \rightarrow \infty$. For the second probability on the right in Equation (3), note that Proposition 4 says that

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbb{P}_{2n} [V'_{2n} \leq 2n(\alpha + \epsilon)] &= \lim_{n \rightarrow \infty} \mathbb{P}_{2n} [V'_{2n} \leq 2n(\alpha + \epsilon)] \\ &= \frac{1}{\pi} \int_0^{\alpha + \epsilon} \frac{1}{\sqrt{x(1-x)}} dx,\end{aligned}$$

and $\epsilon \rightarrow 0$ gives

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \int_0^{\alpha + \epsilon} \frac{1}{\sqrt{x(1-x)}} dx = \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx.$$

Therefore going back to the left hand side of equation (3), we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}_{2n} [V_{2n} < 2n\alpha] \leq \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx.$$

Since $V_{2n} \leq V'_{2n}$, Proposition 4 says that

$$\liminf_{n \rightarrow \infty} \mathbb{P}_{2n} [V_{2n} < 2n\alpha] \geq \frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx.$$

Thus $\lim_{n \rightarrow \infty} \mathbb{P}_{2n} [V_{2n} < 2n\alpha] = \frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx$. To complete, note that

$$\mathbb{P}_{2n} [V_{2n+1} < (2n+1)\alpha] \leq \mathbb{P}_{2n} [V_{2n} < (2n+1)\alpha]$$

and similarly,

$$\mathbb{P}_{2n} [V_{2n+2} < (2n+2)\alpha] \geq \mathbb{P}_{2n} [V_{2n+1} < (2n+2)\alpha].$$

□

Examples and Illustration

Illustration 1

Example. Consider the probability that Heads is in the lead at least 85% of the time:

$$\lim_{n \rightarrow \infty} \mathbb{P}_n [V_n \geq 0.85n] = \lim_{n \rightarrow \infty} (1 - \mathbb{P}_n [V_n < 0.85n]) = 1 - \frac{2}{\pi} \arcsin \sqrt{0.85} = 0.25318.$$

The probability is more than 25%, surprisingly higher than most would anticipate.

Illustration 2

In practice, the formula provides a good approximation even for values of n as small as 20. The table below illustrates the approximation.

Illustration 3

An investment firm sends you an advertisement for their new investment plan. The ad claims that their investment plan, while subject to the “random fluctuations of the market”, yields a net fortune which is on the positive side at least 75% of the time. The company provides a graph of the plan’s outcome to “prove” their claim.

$2n = 20$	k	$p_{2k,2n}$	cdf	k/n	$(2/\pi) \arcsin(\sqrt{\alpha})$
	0	0.1762	0.1762	0	0.0000
	2	0.0927	0.2689	0.1	0.2048
	4	0.0736	0.3426	0.2	0.2952
	6	0.0655	0.4080	0.3	0.3690
	8	0.0617	0.4697	0.4	0.4359
	10	0.0606	0.5303	0.5	0.5000
	12	0.0617	0.5920	0.6	0.5641
	14	0.0655	0.6574	0.7	0.6310
	16	0.0736	0.7311	0.8	0.7048
	18	0.0927	0.8238	0.9	0.7952
	20	0.1762	1.0000	1	1.0000
$2n = 40$	$2k$				
	0	0.1254	0.1254	0	0.0000
	2	0.0643	0.1897	0.05	0.1436
	4	0.0495	0.2392	0.1	0.2048
	6	0.0424	0.2816	0.15	0.2532
	8	0.0383	0.3199	0.2	0.2952
	10	0.0356	0.3555	0.25	0.3333

Figure 2: A comparison of exact and arcsine probability distributions

However, you should be suspicious. Even under the simple null hypothesis that their investment plan will yield a gain of 1 unit with probability 1/2 and will lose a unit with probability 1/2, the arcsine law tells us that the resulting fortune would spend 75% to 100% of its time on the positive side with probability:

$$\frac{2}{\pi} \arcsin(\sqrt{1}) - \frac{2}{\pi} \arcsin(\sqrt{0.75}) = 0.33333$$

That is, “just by chance” the seemingly impressive result could occur about 1/3 of the time. Not enough evidence has been provided to convince us of the claim!

History and Comments

The Arcsine Law was first proved by P. Lévy in 1939 for Brownian motion. Then Erdős and Kac proved the Arcsine Law in 1947 for sums of independent random variables using an Invariance Principle. In 1954 Sparre Andersen proved the Arcsine Law with a combinatorial argument. There are several other ways to prove the Arcsine Law, which means that the Arcsine Law has a surprising variety of proofs.

Sources

This section is adapted from: This section is adapted from: *Heads or Tails*, by Emmanuel Lesigne, Student Mathematical Library Volume 28, American Mathematical Society, [2]. Providence, 2005, Chapter 10.4. Some ideas are adapted from Chapter XIV of the classic text by Feller, [1].



Algorithms, Scripts, Simulations

Algorithm

ArcsineLaw-Simulation

- Comment Post: Empirical probability of random walks being positive at most $100\alpha\%$.
Comment Post: Theoretical Arcsine Law probability $\frac{2}{\pi} \arcsin(\sqrt{\alpha})$
- 1 Set probability of success p
 - 2 Set length of random walk n
 - 3 Set number of trials k
 - 4 Set Arcsine Law parameter α
 - 5 Initialize and fill $k \times n$ matrix of random walks
 - 6 Use vectorization to find where each walk is positive
 - 7 Use vectorization to sum the Boolean vector, *findposwalks*
 - 8 Count how many walks are positive on $< n\alpha$ of the steps, *longleads*
 - 9 **return** Empirical probability *longleads/k*
 - 10 **return** Theoretical probability $\frac{2}{\pi} \arcsin(\sqrt{\alpha})$

Scripts

Scripts

R R script for the Arcsine Law

```
p <- 0.5
n <- 100
k <- 200

alpha <- 0.85

walks <- matrix(0, nrow = k, ncol = n + 1)
rw <- t(apply(2 * matrix((runif(n * k) <= p), k, n) - 1, 1, cumsum)
walks[, 1:n + 1] <- rw

findposwalks <- apply(0 + (walks[, 1:n + 1] > 0), 1, sum)
longleads <- sum(0 + (findposwalks < n * alpha))

prob <- longleads/k
theoretical <- (2/pi) * asin(sqrt(alpha))
```

```

cat(sprintf(" Empirical_probability:_%f\n", prob))
cat(sprintf(" Positive_Walks_Theorem_probability:_%f\n", theoretical

```

Octave Octave script for Arcsine Law

```

p = 0.5;
n = 100;
k = 200;

alpha = 0.85;

walks = zeros(k, n+1);
walks(:, 2:n+1) = cumsum((2 * (rand(k,n) <= p) - 1), 2);

findposwalks = sum( walks(:, 2:n+1) > 0, 2);
longleads = sum( findposwalks < n*alpha );

prob = longleads/k;
theoretical = (2/pi)*asin(sqrt(alpha));

disp(" Empirical_probability:"), disp( prob )
disp(" Arcsine_Law_probability:"), disp( theoretical )

```

Perl Perl PDL script for the Arcsine Law.

```

use PDL::NiceSlice;

$p = 0.5;
$n = 100;
$k = 200;

$alpha = 0.85;

$walks = zeros( $n + 1, $k );
$rw = cumusumover( 2 * ( random( $n, $k ) <= $p ) - 1 );
$walks ( 1 : $n, 0 : $k - 1 ) .= $rw;

$findposwalks = sumover( $walks ( 1 : $n, 0 : $k - 1 ) > 0 );

```



```

$poswalks = sum( $findposwalks < $n * $alpha );

$prob = $poswalks / $k;

use PDL::Constants qw(PI);
use PDL::Math;
$theoretical = ( 2. / PI ) * asin( sqrt($alpha) );

print "Empirical_probability", $prob,
"\n";
print "Positive_Walks_Theorem_probability", $theoretical, "\n";

```

SciPy Scientific Python script for the Arcsine Law.

```

import scipy

p = 0.5
n = 100
k = 200

alpha = 0.85

walks = scipy.zeros((k, n + 1), dtype=int)
rw = scipy.cumsum(2 * (scipy.random.random((k, n)) <= p) - 1, axis=
walks[:, 1:n + 1] = rw

findposwalks = 0 + (walks[:, 1:n + 1] > 0)
poswalks = scipy.sum(scipy.sum(findposwalks, axis=1) < n * alpha)

prob = float(poswalks) / float(k)
theoretical = 2.0 / scipy.pi * scipy.arcsin(scipy.sqrt(alpha))

print 'Empirical_probability:', prob
print 'Positive_Walks_Theorem_probability:', theoretical

```



Problems to Work for Understanding

1. Show that

$$\frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin \sqrt{\alpha}.$$

2. Adapt the scripts with a larger number of trials and longer walks to create an empirical histogram of the Arcsine Law, comparing it with the theoretical density as in Figure 1. Use an increased number of histogram intervals to create a finer representation.



Reading Suggestion:

References

- [1] William Feller. *An Introduction to Probability Theory and Its Applications, Volume I*, volume I. John Wiley and Sons, third edition, 1973. QA 273 F3712.
- [2] Emmanuel Lesigne. *Heads or Tails: An Introduction to Limit Theorems in Probability*, volume 28 of *Student Mathematical Library*. American Mathematical Society, 2005.



Outside Readings and Links:

I check all the information on each page for correctness and typographical errors. Nevertheless, some errors may occur and I would be grateful if you would alert me to such errors. I make every reasonable effort to present current and accurate information for public use, however I do not guarantee the accuracy or timeliness of information on this website. Your use of the information from this website is strictly voluntary and at your risk.

I have checked the links to external sites for usefulness. Links to external websites are provided as a convenience. I do not endorse, control, monitor, or guarantee the information contained in any external website. I don't guarantee that the links are active at all times. Use the links here with the same caution as you would all information on the Internet. This website reflects the thoughts, interests and opinions of its author. They do not explicitly represent official positions or policies of my employer.

Information on this website is subject to change without notice.

Steve Dunbar's Home Page, <http://www.math.unl.edu/~sdunbar1>

Email to Steve Dunbar, `sdunbar1` at `unl dot edu`

Last modified: Processed from L^AT_EX source on March 10, 2015