SCHOOL *of* GRADUATE STUDIES
EAST TENNESSEE STATE UNIVERSITY

East Tennessee State University
Digital Commons @ East Tennessee State University

Electronic Theses and Dissertations

Student Works

5-2019

Generalizations of the Arcsine Distribution

Rebecca Rasnick
East Tennessee State University

Follow this and additional works at: https://dc.etsu.edu/etd

Part of the Discrete Mathematics and Combinatorics Commons, Probability Commons, and the Statistical Theory Commons

Recommended Citation
Rasnick, Rebecca, "Generalizations of the Arcsine Distribution" (2019). *Electronic Theses and Dissertations.* Paper 3565. https://dc.etsu.edu/etd/3565

This Thesis - unrestricted is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.

Generalizations of the Arcsine Distribution

A thesis

presented to

the faculty of the Department of Mathematics and Statistics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

Rebecca Rasnick

May 2019

Anant Godbole, Ph.D. , Chair

Robert Price, Ph.D.

Nicole Lewis, Ph.D.

JeanMarie Hendrickson, Ph.D.

Keywords: arcsine distribution, catalan numbers, random walk, first return to origin

ABSTRACT

Generalizations of the Arcsine Distribution

by

Rebecca Rasnick

The arcsine distribution looks at the fraction of time one player is winning in a fair coin toss game and has been studied for over a hundred years. There has been little further work on how the distribution changes when the coin tosses are not fair or when a player has already won the initial coin tosses or, equivalently, starts with a lead. This thesis will first cover a proof of the arcsine distribution. Then, we explore how the distribution changes when the coin toss is unfair. Finally, we will explore the distribution when one person has won the first few flips.

Copyright by Rebecca Rasnick 2019

All Rights Reserved

ACKNOWLEDGMENTS

I would like to thank Dr. Anant Godbole for letting me part of this amazing project. I would also like to thank my committee and my professors that answered my questions and got me to this point.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The arcsine distribution is a special case of the beta distribution, specifically $Beta(\frac{1}{2}, \frac{1}{2})$. The beta distribution is a continuous probability distribution that is defined on the closed interval $[0, 1]$. The arcsine distribution is a symmetric distribution with a minimum at $\frac{1}{2}$ and vertical asymptotes at 0 and 1. The distribution is normally described as arising from a coin flipping game between two-players typically named Peter and Paul. In this game we assume that the coin is perfectly fair and the flips are independent. Independence, in this case, means the result of the previous flips have no affect on future flips. We will let Peter be player 1 and Paul be player 2. If the coin lands on heads, Peter will get a dollar, and if it lands on tails, Paul will get a dollar. In this paper we will refer to a positive lead as the time spent by the quantity above the x-axis, which is the amount of time Peter is in the lead, in other words the proportion of time there are more heads flipped than tails. The arcsine distribution is also used in Bayesian Statistics. For a coin toss game, the arcsine distribution is the Jeffreys prior. Jeffreys prior is a distribution one uses when there is no previous assumed distribution.

The arcsine distribution is the proportion of time spent above or below the x-axis on a random walk as $n \to \infty$. We consider a random walk to be a broken line segment where a coin is flipped: if Peter wins the line segment will go up one unit and if Paul wins it will go down one unit. The law of long leads states that if we let the coin flipping game go on indefinitely, a lead will form, and since the coin is fair, the broken line segment will spend most of its time on either the positive side or negative side of the x-axis. This results in the bowl shaped distribution demonstrated in Figure 1.

The arcsine distribution has a set of two assumptions that we are going to violate in

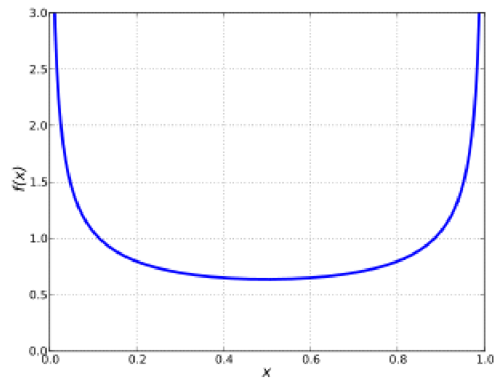

Figure 1: Probability Mass Function of the Arcsine Distribution

this paper. First, is the assumption that the coin is fair. By this we mean there is a equal chance of winning and losing the coin toss. Takács found a generic distribution for when the coin on unfair [13]. Second there is an assumption of there being no initial lead. The distribution when this assumption is violated will be covered in this thesis.

While the arcsine distribution is not a relatively new distribution there have been very little papers trying to find formulas where one of the assumptions is not meet. While we only cover two broken assumption distributions there does exist others. In this thesis we will cover generalizations of the arcsine distribution. The proof of the arcsine distribution is covered in Chapter 2. Chapter 3 will include graphs and R-code that will be useful for visualizing the later information. In Chapter 4 a tractable form of Takács' work will be obtained. Then in Chapter 5 we will cover the work done so far on the distribution for the initial lead. Chapter 6 will deal with potential applications. Finally, Chapter 7 will be the conclusions and possible future work for this topic.

2 PROOF OF THE ARCSINE DISTRIBUTION

To prove the arcsine distribution we will mimic the proof by Shlomo Sternberg [12]. Let us assume that we have we have a fair game of coin tossing where the next flip of the coin is independent of the previous flips and both heads and tails have equal probabilities of happening. We will call player 1 Peter and player 2 Paul. We will denote the successes and failure of Peter as S_0 where $S_0 = 0, S_1, S_2, \ldots$. We represent wins and losses with a time series plot. Each flip will move the particle over one position in the positive x-direction. When Peter wins, the particle will move in the positive y-direction 1 position. When Paul wins, the particle will move one direction in the negative y-direction. This game will continue for $2n$ flips.

Let $k \leq n$. We will denote the probability that the last visit to the origin (when the game is tied) happens at $2k$ as $\alpha_{2k,2n}$. Let

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}$$

be the probability that out of the first 2ν flips exactly ν was where Peter won and the remainder is where Paul won. Essentially 2ν will be the time until a return to the origin. One can approximate $u_{2\nu}$ using Stirling's approximation. Stirling's approximation is

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

This approximation will get closer and closer as n gets larger. Solving we get

$$\begin{aligned}
u_{2\nu} &= 2^{-2\nu}\frac{(2\nu)!}{(\nu!)^2} \\
&\sim 2^{-2\nu}\frac{\sqrt{2\pi}(2\nu)^{2\nu+\frac{1}{2}}e^{-2\nu}}{2\pi\nu^{2\nu+1}e^{-2\nu}} \\
&= \frac{1}{\sqrt{\pi\nu}}.
\end{aligned}$$

We first want to show that we have

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k} \tag{1}$$

and this formula can be approximated by

$$\alpha_{2k,2n} \sim \frac{1}{\pi\sqrt{k(n-k)}}. \tag{2}$$

Lemma 2.1 (The Reflection Principle) *Let A and B be two points in the first Quadrant where $A = (a, \alpha)$ and $B = (b, \beta)$. Let $0 \leq a < b$ where both α and β are positive whole numbers. The number of paths that touch the t-axis between the points A and B is the also number of paths that go from A' to B.*

Proof: Let t be be a point on any path from A to B where t is the first point to touch the t-axis. Let the path from A to T, where $T = (t, 0)$ be reflected across the t-axis. This reflected path will be the path A' to T. This will also give the path from A' to B. This process will assign any path from A to B that has a point that touches the t-axis a path from A' to B. This process is also bijective because any path from A' to B has to cross the t-axis. If one reflects the path from A' to T one gets a path that touches the t-axis and goes from A to B. This is the reflection principle. $\square$

Let there be a path n days long that joins $(0,0)$ to (n,x) where there are p days that have a slope of +1 (days Paul wins) and q days with a slope of -1 (days Peter wins). This means that

$$n = p + q, \quad x = p - q.$$

The number of paths is the number of ways to choose the points where there are p steps that join $(0,0)$ to (n,x) is

$$N_{n,x} = \binom{p+q}{p} = \binom{n}{\frac{n+x}{2}}.$$

If there are no paths that join $(0,0)$ to (n,x) then $N_{n,x} = 0$.

Lemma 2.2 (The Ballot Theorem) *Let n and x be positive numbers. There are*

$$\frac{x}{n} N_{n,x}$$

number of paths that that are strictly about the t-axis for $t > 0$ that join $(0,0)$ to (n,x).

Proof: Since the paths are strictly above the t-axis this means that there are as many paths which join $(0,0)$ to (n,x) as there are paths joining $(1,1)$ to (n,x) that do not touch or cross the t-axis. This is the number of paths join $(0,0)$ to (n,x) minus the number of paths that touch or cross the t-axis. The number of paths that touch or cross the t-axis is the number of paths that join $(1,-1)$ to (n,x) by the reflection principle. $N_{n-1,x-1}$ is the number of paths that join $(1,1)$ to (n,x), and $N_{n-1,x+1}$ is the number of paths that join $(1,-1)$ to (n,x). The number of paths strictly above the t-axis for $t > 0$ that join $(0,0)$ to (n,x) is

$$
\begin{aligned}
N_{n-1,x-1} - N_{n-1,x+1} &= \binom{n-1}{\frac{(n-1)+(x-1)}{2}} - \binom{n-1}{\frac{(n-1)+(x+1)}{2}} \\
&= \binom{n-1}{\frac{n+x+2}{2}} - \binom{n-1}{\frac{n+x}{2}} \\
&= \binom{p+q-1}{p-1} - \binom{p+q-1}{p} \\
&= \frac{(p+q-1)!}{(p-1)!q!} - \frac{(p+q-1)!}{p!(q-1)!} \\
&= \frac{p}{p}\frac{(p+q-1)!}{(p-1)!q!} - \frac{q}{q}\frac{(p+q-1)!}{p!(q-1)!} \\
&= \frac{(p+q-1)!(p-q)}{p!q!} \\
&= \frac{(p+q-1)!(p-q)}{p!q!}\frac{p+q}{p+q} \\
&= \frac{(p+q)!(p-q)}{p!q!(p-q)} \\
&= \frac{p-q}{p+q}\frac{(p+q)!}{p!q!} \\
&= \frac{x}{n}N_{n,x}.
\end{aligned}
$$

□

The previous lemma is called the Ballot Theorem because the Ballot Theorem states that if candidate A gets p votes and candidate B gets q votes where the chance of a vote being p is $\frac{1}{2}$, and candidate A wins $(p > q)$, then the probability that there are more votes for candidate A than B throughout the counting is

$$\frac{p-q}{p+q} = \frac{x}{n}.$$

Let

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

where u_{2n} is the probability that exactly n out of $2n$ steps were positive and the rest negative. This means that u_{2n} is the probability that on the last coin flipping day there would have been as many heads as tails.

Lemma 2.3 *The probability that from day one until the end (day 2n) Paul will be in the lead is* $\frac{1}{2}u_{2n}$. *Which can be written as*

$$P\{S_1 > 0, ...S_{2n} > 0\} = \frac{1}{2}u_{2n}$$

Proof: The possible values for S_{2n} ranges from 2 to $2n$. So

$$P\{S_1 > 0, ...S_{2n} > 0\} = \sum_{r=1}^{n} P\{S_1 > 0, ...S_{2n} = 2r\}$$

By the reflection principle in the Ballot Theorem

$$= 2^{-2n} \sum_{r=1}^{n} (N_{2n-1,2r-1} - N_{2n-1,2r+1})$$

$$= 2^{-2n}(N_{2n-1,1} - N_{2n-1,3} + N_{2n-1,3} - N_{2n-1,5} + ...)$$

Canceling values

$$= 2^{-2n}(N_{2n-1,1} - N_{2n-1,2n+1})$$

One cannot get to 2n+1 in 2n-1 steps $N_{2n-1,2n+1} = 0$

$$= 2^{-2n} N_{2n-1,1}$$

$$= \frac{2^{-(2n-1)}}{2} N_{2n-1,1}$$

the chance of ending at (2n-1,1)starting at (0,0) is $2^{-(2n-1)} N_{2n-1,1}$

$$= \frac{1}{2} p_{2n-1,1}$$

one must be at ± 1 at time $2n-1$ for 50% chance of getting to 0

$$= \frac{1}{2} u_{2n}.$$

□

Lemma 2.4 *The probability that a path never touches the t-axis is*

$$P\{S_1 \neq 0, ...S_{2n} \neq 0\} = u_{2n}.$$

Proof: Because the path never touches or crosses the t-axis it can be either all positive or all negative. So by the previous lemma

$$P\{S_1 \neq 0, ...S_{2n} \neq 0\} = u_{2n}.$$

$\square$

Lemma 2.5 *The probability that a path is either above or on the t-axis is* u_{2n}

$$P\{S_1 \geq 0, ...S_{2n} \geq 0\} = u_{2n}.$$

Proof: Let there be a path that is strictly about the t-axis from day 1 on. This path must cross through $(1, 1)$ and stay above a new horizontal line $s = 1$. The chance of going through $(1, 1)$ first is $\frac{1}{2}$ and staying above the new line is $P\{S_1 \geq 0, ...S_{2n-1} \geq 0\}$. Because $2n - 1$ is odd this means that if $S_{2n-1} \geq 0$ then $S_{2n} \geq 0$. So by lemma 2.3 we have

$$\begin{aligned}\frac{1}{2}u_{2n} &= P\{S_1 > 0, ...S_{2n} > 0\} \\ &= \frac{1}{2}P\{S_1 \geq 0, ...S_{2n-1} \geq 0\} \\ &= \frac{1}{2}P\{S_1 \geq 0, ...S_{2n} \geq 0\}.\end{aligned}$$

$\square$

Let

$$u_{2\nu} = \binom{2\nu}{\nu}2^{-2\nu}$$

be the probability that the game is tied at time 2ν. This also means that exactly ν of the first 2ν flips were for Paul and the rest for Peter.

Lemma 2.6 *Let the probability that from 2k+1 to 2n the last visit to the origin be denoted by*

$$\alpha_{2k,2n},$$

where we have

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}$$

Proof: In order for the last visit to the origin to be at time $2k$ this means that

$$S_{2k} = 0$$

also

$$S_j \neq 0, \quad j = 2k+1, ..., 2n.$$

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}$$

For $S_{2k} = 0$ this means that there are $2^{2k}u_{2k}$ ways to choose the first 2k positions. If we let the point $(2k, 0)$ to be the new origin by lemma 2.4 there are $2^{2n-2k}u_{2n-2k}$ ways to choose the last $2n - 2k$ steps in order for $S_j \neq 0, \quad j = 2k+1, ..., 2n$ to be met. If 2^{2n} is multiplied and divided this proves lemma 2.6. □

Theorem 2.7 $\alpha_{2k,2n}$ *is the probability that from 0 to 2n there are 2k units on the positive side and 2n-2k on the negative side. If* $0 < x < 1$ *the chance that the time spent on the positive time is less than* x *goes to* $\sin^{-1}(\sqrt{x})$ *as* $n \to \infty$.

Proof: Let there be a path of $2n$ time units where $b_{2k,2n}$ be the probability that $2k$ units are above the t-axis. We need to prove that

$$b_{2k,2n} = \alpha_{2k,2n}.$$

If $k = n$ then $\alpha_{2k,2n} = u_0 u_{2n} = u_{2n}$ and $b_{2n,2n}$ is the probability that the path is above the t-axis, lemma 2.5 proves this. The probability that the entire path is below the t-axis is $b_{0,2n} = \alpha_{0,2n}$ by symmetry.

Now we need to prove this for $1 \leq k \leq n-1$. In order for this situation to occur this means that there has to be a return to the origin. Let us suppose that this happens at time $2r$. This means that there are two possibilities for the path from the origin to $(2r, 0)$ which is either entirely above or entirely below the t-axis. If it is the case where it is above, then $r \leq k \leq n-1$, and the path after that point has $2k - 2r$ points above the t-axis. The total number of such paths are

$$\frac{1}{2} 2^{2r} f_{2r} 2^{2n-2r} b_{2k-2r,2n-2r}$$

where f_{2r} is the probability that the first return to the origin is at $2r$

$$f_{2r} = P\{S_1 \neq 0, ..., S_{2r-1} \neq 0, S_{2r} = 0\}.$$

If the path up to $2r$ is below the t-axis then the remaining path has $2k$ units above the t-axis. So $n - r \geq k$ and the total number of such paths are

$$\frac{1}{2} 2^{2r} f_{2r} 2^{2n-2r} b_{2k,2n-2r}.$$

So we get

$$b_{2k,2n} = \frac{1}{2} \sum_{r=1}^{k} f_{2r} b_{2k-2r,2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} b_{2k,2n-2r} \quad 1 \leq k \leq n-1.$$

By induction on n we know that $b_{2k,2n} = u_{2k} u_{2n-2k} = \frac{1}{2}$ when $n = 1$. Assuming that result up to $n - 1$, the formula above becomes

$$b_{2k,2n} = \frac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r}.$$

We claim that the probability of returning and the probability of the first return are associated by

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + ... + f_{2n} u_0.$$

If we return to the origin at time $2n$ then the first return has to happen at $2r$ where $2r \leq 2n$ and return to the origin in $2n - 2r$ time units. The sum in the previous equation is over the possible times of the first return. If we substitute the last equation into the first sum the equation becomes u_{2k} and substituting the previous equation into the second term results in u_{2n-2k}. Thus, the equation before the last becomes

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

which is what we wanted to obtain. $\square$

3 SIMULATIONS

The simulations for the arcine distribution are below. N is the overall number of trials aka, number of games "played" and n is how many flips are in each game. Geers is the up and down lines of code. Geers [4] discovered the formula for if the reward for Peter and Paul is different. He refered to this as a drift. Thus if the coin lands on heads Peter will receive $\$1+\epsilon$ and if it lands on tails Paul will receive $\$1-\epsilon$. The graph shown the proportion of time that Peter has more money than Paul. One only needs to change the up value. The varaible, headsPercentage, will create the unfair coin, which the probability that Peter will win. Depending on the value one wants for p heads and tails will need to be changes with the change in significant figures. For Peter to have an initial lead one needs to change the value of w to be how far ahead Peter is in front of Paul at the start of each game. The rest of the values can be left alone or modified for other purposes.

```
N=10000 # number of trials

n=1000 # number of samples from each trial

up=1.0  # Geer's Drift

down=2-up    # Geer's drift

headsPercentage=.5 # probability of heads

w=0            # lead

prob=rep(0,N) # to store the probabilities in

heads<-rep(1,headsPercentage*100)    # number of heads

tails<-rep(0, 100-headsPercentage*100)
```

```
flipspercentage<-c(heads,tails)

for(j in 1:N){      # running these trials N times
  k=sample(flipspercentage, n, replace = T)# obtaining our sample
  x=rep(w,n+1)      # repeating the initial value for x n+1 times.
  y=rep(0,n+1)      # repeating zeros n+1 times
  q=0               # The proportion of person A being in the lead
  for (i in 1:n){   # Making an array that if they won they get a point
    if (k[i]==1){   # If person A won the coin is 1
      x[i+1]=x[i]+up # upon winning Peter (x-array) gets a point at in
       # position k+1
      y[i+1]=y[i]   # If person A wins Paul's value will remain the same at
# position k+1 (y-arrary)
    } else{         # Person B wins
      y[i+1]=y[i]+down # upon winning Paul (y-array) gets a point at in
       # position k+1
      x[i+1]=x[i]   # If Paul wins person A's value will remain the same at
       # position k+1
    }               # End of else statement
  }                 # End of for array for counting the lead
  for (i in 2:n+1){ # Starting from the second position
# We will begin comparing the scores
    if (x[i]>y[i]){ # If person A is in the lead at flip i-1
```

```
      q=q+1          # They get a point
    } else if (x[i]==y[i]){ # If they are tied at flip i-1
      q=q+.5         # Person A gets half a point
    } else {         # If person B is in the lead at time i-1
      q=q            # Person A get no points
    }          # End of else loop for finding Peter's points
    prob[j]=q/n     # Finding the proportion person A is in the lead
     #during trial j
  }                  # End of for loop
}                    # End of Trials
hist(prob)           # Histogram of the probabilities for the trials
```

For comparison in Figure 2 we have the simulation when all of the assumptions of the arcsine distribution is met.

The graph for an initial lead of 2 is simulated in in Figure 3. We can see that we no longer have the nice u-shaped distribution that we had when the assumptions of the arcsine distribution were fulfilled. Notice, even when Peter has a lead it is still more likely that Paul will most of the time than for the proportion of the lead to be even for both. Thus, it appears that we have a j-shaped distribution.

A simulation of the unfair coin is in Figure 4. In this simulation we had that the chance that Peter wins the coin toss is 51%. Notice, just like with the lead we have

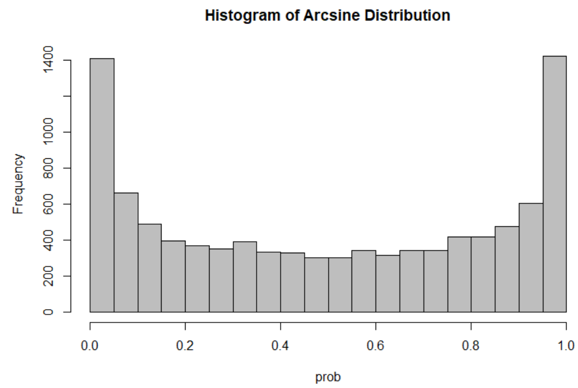

Figure 2: Simulation of the Arcsine Distribution

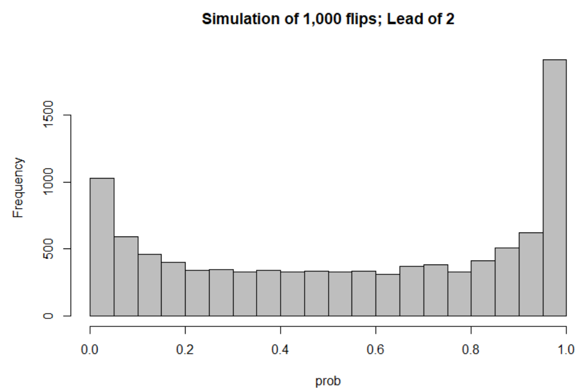

Figure 3: Simulation of the Arcsine Distribution with a lead of 2

a j-shaped distribution.

The final simulated graph is for Geers is found in Figure 5. The drift is 1.01. In other words, if it lands on heads then Peter gets \$1.01 and if it lands on tails Paul will receive \$0.99.

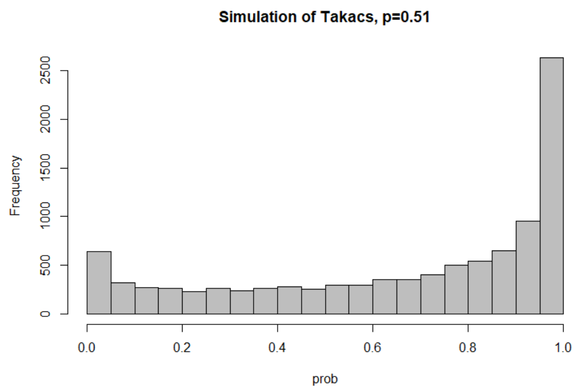

Figure 4: Simulation of Takács' Distribution

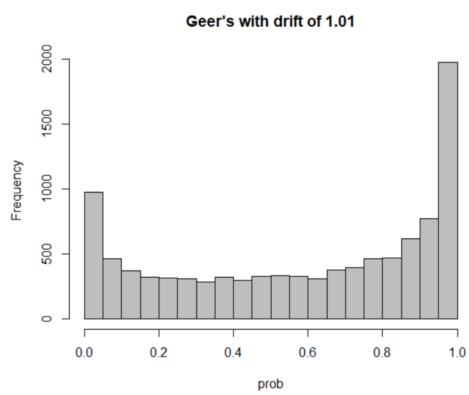

Figure 5: Simulation of Geers' Distribution

4 UNFAIR PROBABILITY VIOLATION

The main focus of this paper is to simplify Takács' equations for the arcsine distribution. He proved that

$$P\{\Delta_n(k) = j\} = P\{\Delta_j = j\}[P\{\rho(k+1) > n-j\} - P\{\rho(k) > n-j\}]. \tag{3}$$

Where $\Delta_n(k)$ is the number of times Peter has a lead of k or more.

We need to simplify this. Also in his paper he proved that

$$P\{\Delta_j = j\} = pP\{\rho(-1) \geq j\} = p - qP\{\rho(1) < j\}. \tag{4}$$

He simplifies equation 3 to

$$\begin{gathered} P\{\Delta_j = j\}[P\{\rho(k+1) > n-j\} - P\{\rho(k) > n-j\}] = P\{\zeta_{n-j} = k\}+ \\ P\{\zeta_{n-j} = k+1\} - \left(1 - \frac{q}{p}\right) P\{\zeta_{n-j} = k+1\} + \left(1 - \frac{p}{q}\right) P\{\zeta_{n-j} < -k-1\}. \end{gathered} \tag{5}$$

In his formula,the values for k are arbitary. Equation 5 is the proportion of time that the random walk is above the value k. Since we are interested in the amount of time Peter is in the lead, let $k = 0$.

Takács further simplified $P\{\Delta_j = j\}$ as

$$P\{\Delta_j = j\} = p - q + q\{P(\zeta_{j-1} = 0) + P(\zeta_{j-1} = -1) + (1 - \frac{p}{q})P(\zeta_{j-1} < -1)\}. \tag{6}$$

Let n be the number of flips and j be time spent above the origin. Let j be odd and $n-j$ be even; the other cases will be similar. Several components will drop out so the probability distribution becomes

$$\begin{aligned} P(\Delta_n(0)=j) =&\{p-q+q*[P(\zeta_{j-1}=0)-(p/(q)-1)+P(\zeta_{j-1}<-1)]\} \\ &\{P(\zeta_{n-j}=0)+(1-\frac{p}{q})P(\zeta_{n-j}<-1)\}. \end{aligned} \tag{7}$$

In equation 7, *binom* is the probability mass function of a binomial distributions and *Binom* is the cumulative mass function. We will denote the probability mass function as pmf and the cummulative mass function as cmf. The probability of success is p and the probability of failure is q.

There exists many inequalities that can bound a binomial cmf however most of these are not good enough for us, so we will use Littlewood's Estimate which was corrected by McKay [10] to solve for Tacáks equation.

For the further calculations notice that we can rewrite Takács distribution with binomial distributions and obtain

$$\begin{aligned} P(\Delta_n(0)=j) =&\{p-q+q*(binom(\frac{j_i-1}{2},j_i-1,p)-(p/(q)-1) \\ &Binom(\frac{j_i-1}{2}-1,j_i-1,p))\}(binom(\frac{(n-j_i)}{2},n-j_i,p)+(1-p/(q)) \\ &*Binom(\frac{n-j_i}{2}-1,n-j_i,p)). \end{aligned} \tag{8}$$

For easier reading we will denote

$$T_1=((p-q)+(q)*(binom(\frac{j_i-1}{2},j_i-1,p)-(p/(q)-1)*Binom(\frac{j_i-1}{2}-1,j_i-1,p)))$$

and

$$T_2 = (binom(\frac{(n-j_i)}{2}, n-j_i, p) + (1-p/(q)) * Binom(\frac{n-j_i}{2} - 1, n-j_i, p)).$$

Let $p = \frac{1}{2} + \frac{c}{\sqrt{n}}$ where $c \in \mathbb{N}$. Note, if $c = 0$ then $p = \frac{1}{2}$ and we have the arcsine distribution. This choice for success and failure probabilities will become apparent later.

We are going to use [10] to simplify Takács formula. Using McKay's notation we have that $q = 1-p$, $\sigma = \sqrt{npq}$, $x = \frac{k-np}{\sigma}$, and $0 \leq E(k;n,p) \leq min\{\sqrt{\frac{\pi}{8}}, \frac{1}{x}\}$. Recall the fomulas for the pmf for binomial is $b(k) = b(k;n,p) = \binom{n}{k} p^k q^{n-k}$ and the binomial cdf is denoted as $B(k;n,p) = \sum_{j=k}^{n} b(j;n,p)$. In McKay's Theorem 2 he states

$$B(k,n,p) = \sigma b(k-1, n-1, p) Y(x) exp\{E(k;n,p)/\sigma\}. \tag{9}$$

We want to simplify equation 7 and get it into a limiting form.

Let us first simplify the pmfs. According to [13] and [10] $P(\zeta_{j-1} = 0) = b(\frac{j-1}{2}, j-1, p)$ and similarly $P(\zeta_{n-1} = 0) = b(\frac{n-j}{2}, n-j, p)$.

Now let us simplify $P(\zeta_{j-1} < -1)$.

$$\begin{aligned} P(\zeta_{j-1} < -1) &= P(\zeta_{j-1} \leq -2) \\ &= Binom(\frac{j-1}{2} - 1, j-1, p) \text{ by [13]} \\ &= B(\frac{j-1}{2} + 1, j-1, q) \text{ since [10] looks at right tailed probabilities} \end{aligned} \tag{10}$$

$$
\begin{aligned}
P(\zeta_{n-j} < -1) &= P(\zeta_{n-j} \leq -2) \\
&= Binom(\frac{n-j}{2} - 1, n-j, p) \text{ by [13]} \\
&= B(\frac{n-j}{2} + 1, n-j, q) \text{ since [10] looks at right tailed probabilities}
\end{aligned} \tag{11}
$$

McKay donotes $k = \frac{j-1}{2} + 1$, $n = j - 1$, and $p = q$, or equivalently, successes and failures are flipped.

Let us now evaluate some of the individual components. We want to first simplify $B(\frac{j-1}{2} - 1, j-2, p)$

$$
\sigma = \sqrt{npq} = \sqrt{(j-1)(\frac{1}{2} - \frac{c}{\sqrt{n}})(\frac{1}{2} + \frac{c}{\sqrt{n}})} \approx \frac{\sqrt{j}}{2}. \tag{12}
$$

The above approximation is reasonable when we take a limit as we will be setting $y = \frac{j}{n}$ thus $0 \leq y \leq 1$.

Now let us solve for $b(\frac{j-1}{2}-1, j-2, p)$

$$
\begin{aligned}
b(\frac{j-1}{2}-1, j-2, p) &= \binom{j-2}{\frac{j-1}{2}-1} p^{\frac{j-1}{2}-1} q^{\frac{j-1}{2}} \\
&\approx \binom{j}{\frac{j}{2}} \left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)^{\frac{j-1}{2}-1} \left(\frac{1}{2}+\frac{c}{\sqrt{n}}\right)^{\frac{j-1}{2}} \\
&\approx 4\binom{j}{\frac{j}{2}} \left(\frac{1}{2}(1-\frac{2c}{\sqrt{n}})\right)^{\frac{j}{2}} \left(\frac{1}{2}(1+\frac{2c}{\sqrt{n}})\right)^{\frac{j}{2}} \\
&\approx 4\binom{j}{\frac{j}{2}} \left(\frac{1}{2}\right)^{j} \left(1-\frac{4c^2}{n}\right)^{\frac{j}{2}} \\
&\approx 4\sqrt{\frac{2}{\pi j}} e^{\frac{-2c^2 j}{n}} \\
&\text{by Stirling's approximation and the estimate } (1-x) \approx e^{-x}
\end{aligned} \tag{13}
$$

and $b(\frac{j-1}{2}, j-1, p)$

$$
\begin{aligned}
b(\frac{j-1}{2}, j-1, p) &= \binom{j-1}{\frac{j-1}{2}} p^{\frac{j-1}{2}-1} q^{\frac{j-1}{2}} \\
&\approx \binom{j}{\frac{j}{2}} \left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)^{\frac{j-1}{2}} \left(\frac{1}{2}+\frac{c}{\sqrt{n}}\right)^{\frac{j-1}{2}} \\
&\approx 2\binom{j}{\frac{j}{2}} \left(\frac{1}{2}(1-\frac{2c}{\sqrt{n}})\right)^{\frac{j}{2}} \left(\frac{1}{2}(1+\frac{2c}{\sqrt{n}})\right)^{\frac{j}{2}} \\
&\approx 2\binom{j}{\frac{j}{2}} \left(\frac{1}{2}\right)^{j} \left(1-\frac{4c^2}{n}\right)^{\frac{j}{2}} \\
&\approx 2\sqrt{\frac{2}{\pi j}} e^{\frac{-2c^2 j}{n}} \\
&\text{by Stirling's approximation and the estimate } (1-x) \approx e^{-x}
\end{aligned} \tag{14}
$$

We now need to solve for x.

$$\begin{aligned} x &= \frac{k - np}{\sigma} \\ &= \frac{\frac{j-1}{2} - (j-1)(\frac{1}{2} - \frac{c}{\sqrt{n}})}{\frac{\sqrt{j}}{2}} \\ &= \frac{2(j-1)c}{\sqrt{nj}} \\ &\approx 2c\sqrt{\frac{j}{n}} \\ &= 2c\sqrt{y}. \end{aligned} \tag{15}$$

Thus x varies from 0 to $2c$. So $Y(x) = Y(2c\sqrt{y})$.

Now we need to solve for $Y(x)$. According to [10]

$$Y(x) = \frac{\int_x^\infty \phi(u)du}{\phi(x)}, \tag{16}$$

where $\phi(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$.

The figure of $Y(x)$ can be seen below in figure 6

Now let us look at $e^{\frac{E(k,n,p)}{\sigma}}$. Note, $e^{\frac{0 \le E(k;n,p) \le min\{\sqrt{\frac{\pi}{8}}, \frac{1}{x}\}}{\sigma}}$ so as $c \to \infty$ we get that

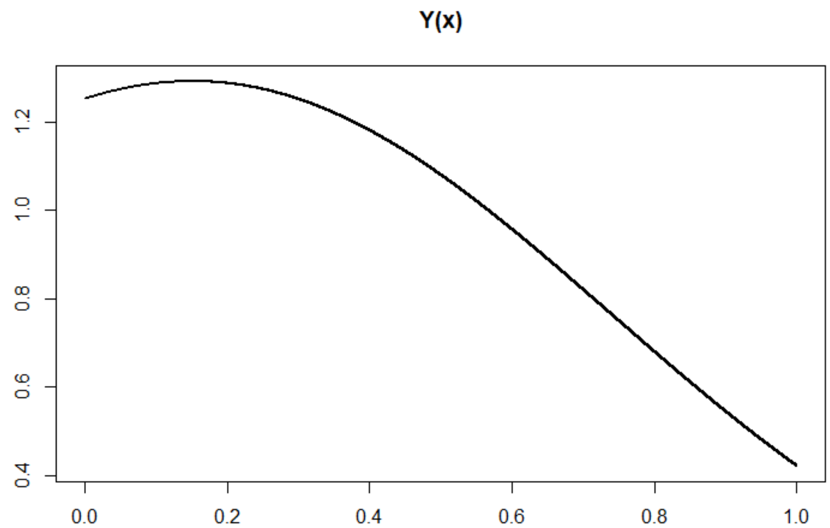

Figure 6: Graph of $Y(x)$

$e^{\frac{E(k,n,p)}{\sigma}}$ is a bounded function. So now we can simplify T_1

$$
\begin{aligned}
T_1 &= (p-q)+(q)*(binom(\frac{j-1}{2}, j-1, p)-(p/(q)-1)*Binom(\frac{j_i-1}{2}-1, j-1, p)) \\
&= \frac{2c}{\sqrt{n}}+\left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)\left\{binom(\frac{j-1}{2}, j-1, p)-\left(\frac{4c}{\sqrt{n}}\right)*Binom(\frac{j_i-1}{2}-1, j_i-1, p)\right\} \\
&= \frac{2c}{\sqrt{n}}+\left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)\{binom(\frac{j-1}{2}, j-1, p)-\left(\frac{4c}{\sqrt{n}}\right) \\
&\times \sigma binom(\frac{j-1}{2}-1, j-2, p)Y(2c\sqrt{y})exp\{\frac{E}{\sigma}\}\} \\
&= \frac{2c}{\sqrt{n}}+\left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)\left\{2\sqrt{\frac{2}{\pi j}}e^{\frac{-2c^2j}{n}}-\left(\frac{4c}{\sqrt{n}}\right)\frac{\sqrt{j}}{2}4\sqrt{\frac{2}{\pi j}}e^{\frac{-2c^2j}{n}}Y(2c\sqrt{y})exp\{\frac{2E}{\sqrt{j}}\}\right\} \\
&= \frac{2c}{\sqrt{n}}+\left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)\left(2\sqrt{\frac{2}{\pi j}}e^{-2c^2y}\right)\left(1-\frac{4c}{\sqrt{n}}Y(2c\sqrt{y})e^{\frac{2E}{\sqrt{j}}}\right).
\end{aligned} \tag{17}
$$

Now let us simplify $B(\frac{n-j}{2}+1, n-j, q)$. We need to find σ, $b(\frac{n-j}{2}-1, n-j, p)$,

x, and $Y(x)$.

Let us first solve for $b(\frac{n-j}{2}, n-j, p)$

$$
\begin{aligned}
b\left(\frac{n-j}{2}, n-j, p\right) &= \binom{n-j}{\frac{n-j}{2}} p^{\frac{n-j}{2}} q^{\frac{n-j}{2}} \\
&= \binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2} + \frac{c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \left(\frac{1}{2} - \frac{c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \\
&= \binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}\left(1 - \frac{2c}{\sqrt{n}}\right)\right)^{\frac{n-j}{2}} \left(\frac{1}{2}\left(1 + \frac{2c}{\sqrt{n}}\right)\right)^{\frac{n-j}{2}} \\
&= \binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}\right)^{n-j} \left(1 - \frac{2c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \left(1 + \frac{2c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \\
&= \binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}\right)^{n-j} \left(1 - \frac{4c^2}{n}\right)^{\frac{n-j}{2}} \\
&= \frac{1}{\sqrt{2\pi(n-j)}} 2^{n-j} \left(\frac{1}{2}\right)^{n-j} \left(1 - \frac{4c^2}{n}\right)^{\frac{n-j}{2}} \\
&= \frac{1}{\sqrt{2\pi(n-j)}} exp\left\{\frac{-2c^2(n-j)}{n}\right\} \\
&= \frac{1}{\sqrt{2\pi(n-j)}} exp\left\{-2c^2(1-y)\right\}.
\end{aligned} \tag{18}
$$

Now we solve for $b\left(\frac{n-j}{2}-1, n-j-1, p\right)$

$$
\begin{aligned}
b\left(\frac{n-j}{2}-1, n-j-1, p\right) &= \binom{n-j-1}{\frac{n-j}{2}-1} p^{\frac{n-j}{2}} q^{\frac{n-j}{2}-1} \\
&\approx \binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}+\frac{c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)^{\frac{n-j}{2}-1} \\
&\approx 2\binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}\left(1-\frac{2c}{\sqrt{n}}\right)\right)^{\frac{n-j}{2}} \left(\frac{1}{2}\left(1+\frac{2c}{\sqrt{n}}\right)\right)^{\frac{n-j}{2}} \\
&= 2\binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}\right)^{n-j} \left(1-\frac{2c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \left(1+\frac{2c}{\sqrt{n}}\right)^{\frac{n-j}{2}} \\
&= 2\binom{n-j}{\frac{n-j}{2}} \left(\frac{1}{2}\right)^{n-j} \left(1-\frac{4c^2}{n}\right)^{\frac{n-j}{2}} \\
&= 2\frac{1}{\sqrt{2\pi(n-j)}} 2^{n-j} \left(\frac{1}{2}\right)^{n-j} \left(1-\frac{4c^2}{n}\right)^{\frac{n-j}{2}} \\
&= \frac{2}{\sqrt{2\pi(n-j)}} exp\left\{\frac{-2c^2(n-j)}{n}\right\} \\
&= \frac{2}{\sqrt{2\pi(n-j)}} exp\left\{-2c^2(1-y)\right\}.
\end{aligned}
\tag{19}
$$

Now let us solve for σ for T_2

$$
\sigma = \sqrt{npq} = \sqrt{(n-j)\left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)\left(\frac{1}{2}+\frac{c}{\sqrt{n}}\right)} \approx \frac{\sqrt{n-j}}{2}. \tag{20}
$$

Next, let us solve

$$
\begin{aligned}
x = \frac{k-np}{\sigma} &= \frac{\frac{n-j}{2}-(n-j)\left(\frac{1}{2}-\frac{c}{\sqrt{n}}\right)}{\frac{\sqrt{n-j}}{2}} \\
&= \frac{\frac{c}{\sqrt{n}}(n-j)^{\frac{1}{2}}}{\frac{\sqrt{n-j}}{2}} = \frac{2c\sqrt{n-j}}{\sqrt{n}} \\
&= 2c\sqrt{1-y}.
\end{aligned}
\tag{21}
$$

Now for we know for T_2 our x also varies from 0 to $2c$.

Thus T_2 can be simplified to be

$$\begin{aligned}
T_2 &= binom\left(\frac{n-j}{2}, n-j, p\right) + (1-\frac{p}{q})Binom\left(\frac{n-j}{2}, n-j, p\right) \\
&= binom\left(\frac{n-j}{2}, n-j, p\right) + \frac{4c}{\sqrt{n}}\sigma binom\left(\frac{n-j}{2} - 1, n-j-1, p\right) Y(x)e^{\frac{E}{\sigma}} \\
&= \frac{1}{\sqrt{2\pi(n-j)}}e^{-2c^2(1-y)} + \frac{4c}{\sqrt{n}}\frac{\sqrt{n-j}}{2}\frac{2}{\sqrt{2\pi(n-j)}}e^{-2c^2(1-y)}Y(2c\sqrt{1-y})e^{\frac{2E}{\sqrt{n-j}}} \\
&= \frac{1}{\sqrt{2\pi}}e^{-2c^2(1-y)}\left(\frac{1}{\sqrt{n-j}} + \frac{4c}{\sqrt{n}}Y(2c\sqrt{1-y})e^{\frac{2E}{\sqrt{n-j}}}\right).
\end{aligned} \tag{22}$$

For further simplification let us denote

$$T_{11} = \frac{2c}{\sqrt{n}} \tag{23}$$

and

$$T_{12} = \left(\frac{1}{2} - \frac{c}{\sqrt{n}}\right)\left(2\sqrt{\frac{2}{\pi j}}e^{-2c^2 y}\right)\left(1 - \frac{4c}{\sqrt{n}}Y(2c\sqrt{y})e^{\frac{2E}{\sqrt{j}}}\right) \tag{24}$$

Recall, we want to know the resulting equation when we multiple T_1 and T_2 together. So since $T_1 = T_{11} = T_{12}$ then $T_1 \times T_2 = T_{11}T_2 + T_{12}T_2$. Let us solve for $T_{11}T_2$ first.

$$\begin{aligned}
T_{11}T_2 &= \frac{2c}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}e^{-2c^2(1-y)}\left(\frac{1}{\sqrt{n-j}} + \frac{4c}{\sqrt{n}}Y(2c\sqrt{1-y})e^{\frac{2E}{\sqrt{n-j}}}\right). \\
&= \frac{2c}{n}\frac{1}{\sqrt{2\pi}}e^{-2c^2(1-y)}\left(\frac{1}{\sqrt{1-y}} + 4cY(2c\sqrt{1-y})e^{\frac{2E}{\sqrt{n-j}}}\right).
\end{aligned} \tag{25}$$

Now let us solve for $T_{12}T_2$

$$T_{12}T_2 = \left(\frac{1}{2} - \frac{c}{\sqrt{n}}\right)\left(2\sqrt{\frac{2}{\pi j}}e^{-2c^2 y}\right)\left(1 - \frac{4c}{\sqrt{n}}Y(2c\sqrt{y})e^{\frac{2E}{\sqrt{j}}}\right) \times \frac{1}{\sqrt{2\pi}}e^{-2c^2(1-y)}\left(\frac{1}{\sqrt{n-j}} + \frac{4c}{\sqrt{n}}Y(2c\sqrt{1-y})e^{\frac{2E}{\sqrt{n-j}}}\right) \tag{26}$$

$$T_{11}T_2 \approx \frac{2c}{n}\frac{1}{\sqrt{2\pi}}\frac{e^{-2c^2(1-y)}}{\sqrt{1-y}} \tag{27}$$

where $y = \frac{j}{n}$. Which is approximately

$$T_{11}T_2 \approx \sqrt{\frac{2}{\pi}}\frac{ce^{-2c^2(1-y)}}{\sqrt{1-y}} \tag{28}$$

as $n \to \infty$. Also we have

$$T_{12}T_2 \approx \frac{1}{n}\frac{e^{-2c^2}}{\pi\sqrt{\frac{j}{n}\left(\frac{n-j}{n}\right)}}. \tag{29}$$

So the second part of the density is approximately

$$T_{12}T_2 \approx \frac{e^{-2c^2}}{\pi\sqrt{y(1-y}} \tag{30}$$

as $n \to \infty$, and this equals the arcsine distribution if you let $c = 0$. Thus $T_{11}T_2$ is the skewing component and $T_{12}T_2$ is the arcsine component.

Theorem 4.1 *If $c = 0$ and we let $n \to \infty$ we obtain*

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

Proof: Note, the $\frac{1}{n}$ can be thought of like the width of the rectangles formed for a Riemann and the arcsine part, $\frac{1}{\pi\sqrt{y(1-y)}}$ is our height. Just like with Riemann integration we want the width of our rectangles to go to 0, so we let $n \to \infty$. Thus, we obtain the arcsine distribution.

$\square$

Our main result is the distribution when we let $c \neq 0$ and $n \to \infty$.

Theorem 4.2 *The proportion of time Peter spends in the lead if he has success probability of $p = \frac{c}{\sqrt{n}}$ where $c \neq 0$ and $n \to \infty$ then*

$$f(y) \approx \sqrt{\frac{2}{\pi}} c e^{-2c^2(1-y)} \frac{1}{\sqrt{1-y}} + \sqrt{\frac{2}{\pi}} e^{-2c^2(1-y)} \frac{1}{\sqrt{y(1-y)}}$$

For discussion we will call $\frac{2c}{\sqrt{2\pi}} e^{-2c^2(1-y)} \frac{1}{\sqrt{1-y}}$ $T_{11}T_2$ and $\sqrt{\frac{2}{\pi}} e^{-2c^2(1-y)} \frac{1}{\sqrt{y(1-y)}}$ is $T_{12}T_2$. Notice,$T_{12}T_2$ is the arcsine distribution with some constants. $T_{11}T_2$ is the skewing component. Note, as c increases the $T_{11}T_2$ becomes the dominate term. The skewing component is in figure 7, the arcsine component is in figure 8, and our final distribution can be seen in figure 9.

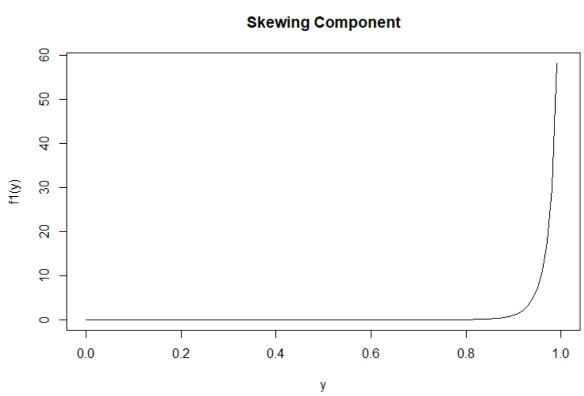

Figure 7: Skewing Component

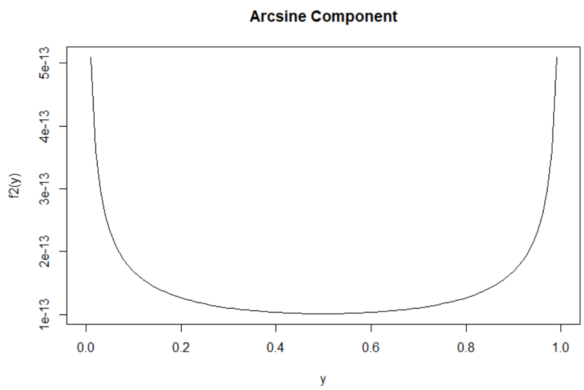

Figure 8: Arcsine Component

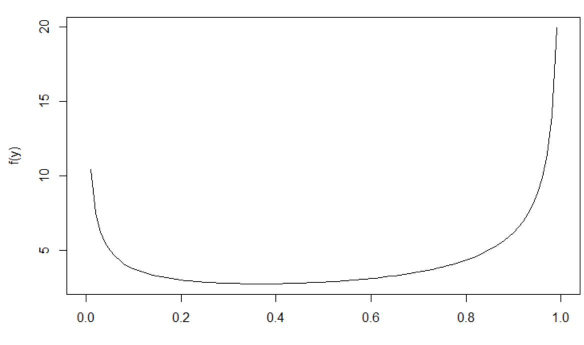

Figure 9: Graph of $f(y)$

5 INITIAL LEAD VIOLATION

In this chapter we will explore the distribution when Peter begins the game with a few dollars initially. In other words, Peter will have a lead in the coin toss game. So now we are interested in how the distribution changes with an initial lead.

Before we explore the distribution let us discuss a famous paradox called St. Petersburg Paradox. According to [9] the St. Petersburg Paradox is a fair coin flipping game where the player flips the coin until the coin lands on tails. If it takes n flips the player will receive $\$2^n$. So if the first flip results in tails then they get \$2. Note, the probability of receiving \$2 is $\frac{1}{2}$. Also the probability that the player receives $\$2^2 = \$4$ is $\frac{1}{4}$. This is true for all n so the chance that you get $\$2^n$ is $\frac{1}{2^n}$. Note, when one multiplies the amount they would make by the probability that even happens is 1. Recall, the expected value of a discrete probability function is

$$E(winnings) = \sum_{i=1}^{\infty} P(x_i)f(x_i),$$

where $P(x_i)$ is the probability of event x_i and $f(x_i)$ is the amount they make. Since one has to flip then the lower bound has to be 1 and since we can theoretically never obtain a tail in an infinite number of flip, because the coin flips are independent, our upper bound is ∞. Plugging our variables into the equation we obtain

$$E(winnings) = \sum_{i=1}^{\infty} \frac{1}{2^n}\$2^n = \$1 + \$1 + \cdots = \$\infty.$$

Thus, the expected reward one would receive is $\$\infty$. Hence, if the owner of the event wanted to make as much money as he lost then they would have to charge an infinite amount of money for each customer to expect to break even. This is a paradox because the actual winnings is finite with probability 1.

$$P(\text{winnings are finite}) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1.$$

We will have a similar situation in the initial lead case as will be apparent later. Let m be the initial lead and let n be the number of steps needed for the first return to the origin. Let m be odd, thus n is also odd. The probability that n steps will be needed to reach the origin is [7]

$$P(\text{First hit origin in n steps}) = \frac{m}{n}\binom{n}{\frac{m+n}{2}}\frac{1}{2^n}. \tag{31}$$

To show that this results in a St. Petersburg Paradox situation let $m = 1$ since the time need for any lead will be larger than for $m = 1$.

$$
\begin{aligned}
\sum_{m=1,3,5,\ldots} \frac{1}{n}\binom{n}{\frac{1+n}{2}}\frac{1}{2^n} \times n = & \sum_{m=1,3,5,\ldots} \binom{n}{\frac{1+n}{2}}\frac{1}{2^n} \\
= & \sum_{m=1,3,5,\ldots} \left(\frac{n}{e}\right)^n \frac{\sqrt{2\pi n}}{\sqrt{2\pi\frac{n+1}{2}}} \\
& \left(\frac{2e}{n+1}\right)^{\frac{n+1}{2}} \left(\frac{2e}{n-1}\right)^{\frac{n-1}{2}} \frac{1}{\sqrt{2\pi\frac{n-1}{2}}}\frac{1}{2^n} \\
= & \sum_{m=1,3,5,\ldots} \frac{n^n\sqrt{2\pi n}}{\sqrt{\pi(n+1)}} \left(\frac{1}{n+1}\right)^{\frac{n+1}{2}} \left(\frac{1}{n-1}\right)^{\frac{n-1}{2}} \frac{1}{\sqrt{\pi(n-1)}} \\
= & \sum_{m=1,3,5,\ldots} \frac{\sqrt{2n}}{\sqrt{(n+1)}\sqrt{\pi(n-1)}} \frac{n^n}{(n+1)^{\frac{n+1}{2}}(n-1)^{\frac{n-1}{2}}} \\
= & \sum_{m=1,3,5,\ldots} \frac{\sqrt{2n}}{\sqrt{(n+1)}\sqrt{\pi(n-1)}} \frac{1}{\left(\frac{n+1}{n}\right)^{\frac{n+1}{2}}\left(\frac{n-1}{n}\right)^{\frac{n-1}{2}}} \\
= & \sum_{m=1,3,5,\ldots} \frac{\sqrt{2n}}{\sqrt{(n+1)}\sqrt{\pi(n-1)}} \frac{1}{\left(1+\frac{1}{n}\right)^{\frac{n+1}{2}}\left(1-\frac{1}{n}\right)^{\frac{n-1}{2}}} \\
= & \sum_{m=1,3,5,\ldots} \frac{\sqrt{2n}}{\sqrt{(n+1)}\sqrt{\pi(n-1)}} \frac{1}{e^{\frac{1}{2}}e^{-\frac{1}{2}}} \\
= & \sum_{m=1,3,5,\ldots} \frac{\sqrt{2n}}{\sqrt{(n+1)}\sqrt{\pi(n-1)}} \\
= & \sum_{m=1,3,5,\ldots} \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}} = \infty
\end{aligned}
\tag{32}
$$

Thus, if Peter has a lead he is expected to win. Which is similar to the St. Petersburg Paradox

$$P(\text{finite time needed to hit origin}) = 1.$$

Then why does

$$\sum \frac{m}{n}\binom{n}{\frac{m+n}{2}}\frac{1}{2^n} = 1?$$

The answer is that we have Catalan convolutions. Catalan convolutions are generalizations of Catalan numbers. A Catalan convolution is the number of paths where one goes north and east from $(k, 1)$ to (n, n) that do not cross the $y = x$ line. [1] denoted the Catalan convolution formula as

$$C_{n,k} = \frac{k}{2n-k}\binom{2n-k}{n}.$$

For example, let $k = 1$

$$\frac{1}{2n-1}\binom{2n-1}{n} = \frac{1}{2n-1}\frac{(2n-1)!}{n!(n-1)!} \tag{33}$$

$$= \frac{(2n-2)!}{n!(n-1)!} \tag{34}$$

$$= \frac{(2n-2)!}{n(n-1)!(n-1)!} \tag{35}$$

$$= \frac{1}{n}\binom{2n-2}{n-1} \tag{36}$$

$$= C_{n-1}. \tag{37}$$

What does the Catalan numbers have to do with leads? We want to show that the Catalan numbers are related to the lead formula.

Lemma 5.1 $C_{n,k} = \frac{m}{n}\binom{n}{\frac{m+n}{2}}$.

We want to show that

$$\frac{k}{2n-k}\binom{2n-k}{n} = \frac{m}{n}\binom{n}{\frac{m+n}{2}}.$$

Let $2n - k \to n$ and $k \to m$ this implies that $2n \to n + m$. Thus, $n = \frac{n+m}{2}$. Plugging in these values into the Catalan convolution formula produces our results. $\square$

For our formula we will define N, m, n, i, and j. N is the total number of flips, m is the initial lead, n is the current number of flips, i is the number of every other step taken after m, and j the number of steps until the origin is hit.

$$P(\text{First hit origin in n steps}) = \frac{m}{n}\binom{n}{\frac{m+n}{2}}\frac{1}{2^n} = \pi_i.$$

$$\begin{aligned} P(\text{In } N-j \text{ trials Peter is in the lead } n-j) &= u_{n-j}u_{N-j} \\ &= \binom{2(n-j)}{n-j}2^{-2(n-j)}\binom{2(N-j)}{N-j}2^{-2(N-j)}. \end{aligned}$$

$$P(\text{Peter wins } n) = \sum_{j=m}^{n} \pi_j P(\text{In } N-j \text{ trials Peter is in the lead } n-j),$$

P(Peter wins n)=$\sum_{j=m}^{n} \frac{m}{n}\binom{n}{\frac{m+n}{2}}\frac{1}{2^n}\binom{2(n-j)}{n-j}2^{-2(n-j)}\binom{2(N-j)}{N-j}2^{-2(N-j)}$.

6 POSSIBLE APPLICATIONS

In this thesis we will discuss four possible applications for Takács' distribution as well as the arcsine distribution. First, we will discuss a hypothesis test for a slightly unfair coin, second, a dynamic brain connectivity model, third, a test for a simple random sample, and finally, applications for infectious disease modeling.

A new hypothesis test for a slightly unfair coin needs to be created. Currently if one uses a t-test the power of the test is close to 0 when the unfairness is small. This prevents traditional hypothesis testing from being practical.

Second, an application for a connectivity model was suggested by Dr. Heather Shappell in an email on December 5, 2017. In that email she stated:

> My one thought was maybe something with dynamic brain (perhaps functional) connectivity. I am currently doing work where we have data for a subject for the length of the time they are in an fMRI machine. At every 2 seconds or so, we have brain signals measured at each region of their brain. In the past, researchers would try and construct one brain network (i.e. based on the entire length of the scan) for the subject. This can be done in a variety of ways, but the simplest is to calculate the correlation between each brain region using the time series data. Pairs of regions with a correlation above a certain threshold are assigned an edge.
>
> Nowadays, it's becoming popular to *not* construct just one network using the entire time series. Instead, researchers want to allow the network to vary throughout the scan. The question then becomes, at what time

points does the subject switch to a new network state? And what are the states/unique networks that the subject entered? We think we have an accurate way of estimating that.

So, I was thinking somewhat along those lines. Maybe the arcine distribution could be applied to the simplest case of just assuming two network states. One thing we are interested in is how much time people are spending in each state and whether this predicts disease status or other behavioral traits. So, perhaps we can ask the question, "What is the proportion of time during the scan that the subject had been in state A more times than they have been in state B?" Perhaps this could be compared to the arcsine distribution. It may be the case that the person is most likely to switch which brain state is "in the lead" at the beginning of the scan or at the end of the scan... which I think seems to be what the arcsine distribution says?

Finally, there was two suggestions by Dr. Adam Sima. The first was discussed in an in-person interview on February 15, 2019, and was about making a test to see if data for a survey was truly drawn from a simple random sample or if it could have been another sampling technique. In an email on March 18, 2019, he also suggested we could possibly use the arcsine distribution to identify changes in antibiotic resistance in infectious disease modeling.

7 CONCLUSIONS AND FUTURE WORK

This thesis covered several aspects of the arcsine distribution. We used Simulations, proved the distribution, simplified previous work, and set up a new distribution. There is still much work to be done for the Arcsine distribution.

Some further works include simplifying Geers' formula, explore when independence cannot be assumed, figure out if any of the possible applications are viable, and to simplify the formula for an initial lead.

BIBLIOGRAPHY

[1] Samuel Connolly, Zachary Gabor, and Anant Godbole *The Location of the First Ascent in a 123-Avoiding Permutation* Integers: Electronic Journal of Combinatorial Number Theory 15, Paper # A13, 2015

[2] Steven R. Dunbar *Topics in Probability Theory and Stochastic Processes* University of Nebraska-Lincoln, 2015

[3] William Feller *An Introduction to Probability Theory and Its applications Volume I Third Edition.* Wiley Eastern Private Limited New Delhi 1968

[4] Ernest Schulte-Geers and Wolfgang Stadje *Small Drift Limit Theorems For Random Walks* Bundesamt für Sicherheit in der Informationstechnik (BSI), Godesberger Allee 185 189, 53175 Bonn, Germany

[5] G. R. Grimmett and D. R. Stirzaker *Probability and Random Processes Second Edition* Oxford Science Publications 1992

[6] Christopher Hoffman, Douglas Rizzolo, and Erik Slivken *Fixed Points of 321-Avoiding Permutations* Proc. Amer. Math. Soc. 147, 861-872, 2019

[7] Remco van der Hofstad and Michael Keane *An Elementary Proof of the Hitting Time Theorem*
http://www.win.tue.nl/ rhofstad/monthly753-756-hofstad.pdf The Mathematical Association of America. Monthly 115. October 2008

[8] Aimé Lachel *Sojourn Time in* $\mathbb{Z}^+$ *for the Bernoulli Random Walk on* $\mathbb{Z}$ ESAIM: PS Volume 16, 2012

[9] Robert Martin *The St. Petersburg Paradox* The Stanford Encyclopedia of Philosophy https://plato.stanford.edu/archives/fall2017/entries/paradox-stpetersburg Fall 2017

[10] Brendan McKay *On Littlewood's Estimate for the Binomial Distribution.* Adv. Appl. Prob., 21, 475-478, 1989

[11] Chris H. Rycroft *The Arcsine Distribution* https://ocw.mit.edu/courses/mathematics/ 18-366-random-walks-and-diffusion-fall-2006/lecture-notes/lec1406.pdf 26 October 2006

[12] Shlomo Sternberg. *Lecture 8 The arc sine law. Random walk. The ballot theorem.* http://www.math.harvard.edu/library/sternberg/slides/1180908.pdf Accessed 18 January 2017

[13] Lajos Takács *On a Generalization of the Arc-Sine Law.* Institute of Mathematical Statistics, The Annals of Applied Probability, Vol. 6, No. 3 1035-1040, 1996

VITA

REBECCA RASNICK

Education:	M.S. Mathematical Sciences, East Tennessee State University Johnson City, Tennessee 2019
	B.S. Mathematics (Statistics), East Tennessee State University Johnson City, Tennessee 2017
	A.S. Mathematics, Northeast State Community College, Blountville, Tennessee 2015
Professional Experience:	Teaching Assistant , East Tennessee State University, Johnson City, Tennessee, 2017–2019
	Tutor, East Tennessee State University, Johnson City, Tennessee, 2015–2019
	Tutor, Northeast State Community College, Blountville, Tennessee, 2014–2015