5-2020

An Analysis of the First Passage to the Origin (FPO) Distribution

Aradhana Soni
East Tennessee State University

An Analysis of the First Passage to the Origin (FPO) 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

Aradhana Soni

May 2020

Anant Godbole, Ph.D., Chair

Michele Joyner, Ph.D.

JeanMarie Hendrickson, Ph.D.

Keywords: random walk, first passage to the origin, MLE, Bayesian analysis

ABSTRACT

An Analysis of the First Passage to the Origin (FPO) Distribution

by

Aradhana Soni

What is the probability that in a fair coin toss game (a simple random walk) we go bankrupt in n steps when there is an initial lead of some known or unknown quantity $\$m$? What is the distribution of the number of steps N that it takes for the lead to vanish? This thesis explores some of the features of this first passage to the origin (FPO) distribution. First, we explore the distribution of N when m is known. Next, we compute the maximum likelihood estimators of m for a fixed n and also the posterior distribution of m when we are given that m follows some known prior distribution.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Anant Godbole for letting me be a part of this educationally enriching project and supporting me throughout. I would also like to thank my committee members, Dr. Michele Joyner and Dr. JeanMarie Hendrickson for their continued help.

Thanks to Rebecca Rasnick for motivating me to take up this project. I want to thank my graduate school colleagues Gaffar Solihu and Vincent Onyame for their valuable feedback and support during my thesis.

I would also like to acknowledge the Department of Mathematics and Statistics at East Tennessee State University (ETSU) and all the professors in the department. Heartfelt gratitude to Dr. Teresa Haynes for such an inspiring teaching style and Dr. Nicole Lewis for helping me and being my mentor throughout the graduate school.

Finally, I would like to express my very profound gratitude to my husband for providing me unfailing support and continuous encouragement throughout the process of research and writing thesis. To my parents and sisters for always believing in me, supporting and motivating me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Background and Motivation

The starting point for this thesis is Polya's theorem ([7]) which states that a symmetric random walk in d dimensions that starts at the origin $(0, 0, ..., 0)$ returns to the origin, with probability 1 if $d = 1$ or 2 but has a positive probability of never returning to (0,0,....,0) if $d \geq 3$. For all $d \geq 1$, however (particularly for $d = 1$, 2) the expected return time to the origin is ∞. Let us be more specific for $d = 1$: we have with W being the time of first return to the origin,

$$P(W = 2n) = \frac{\binom{2n}{n}}{(2n-1)2^{2n}}$$

The distribution N that we study is related to this distribution, and the fact that $E(W = \infty)$ follows in much the same way as the proof of the fact that $E(N) = \infty$.

The key idea, however, is that the probability mass function $P(W = 2n)$ is very intractable for large n even though it is straightforward to show (as we do for N) that $P(W = 2n) \sim \frac{C}{n^{\frac{3}{2}}}$. The distribution of N is further complicated by the presence of the parameter $m \to \infty$.

In general for $d = 1$ or 2, it is challenging to find summary statistics like the quantiles of the distribution of W and it is even more difficult in our case.

1.2 Catalan Numbers

The FPO distribution is related to Catalan convolutions [9] which in turn are generalizations of Catalan numbers [11]. For this reason, we start with a description of Catalan numbers and some of the ways in which they arise. There are many

equivalent ways to see how the Catalan numbers, C_n, arise.

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

See Stanley's book [11] for at least 100 other contexts. We choose as our basic definition their historically first combinatorial interpretation. Let P_{n+2} denote a convex polygon in the plane with $n+2$ vertices (or convex $(n+2)$-gon). A triangulation of P_{n+2} is a set of $n-1$ diagonals of P_{n+2} which do not cross in their interiors. It follows easily that these diagonals partition the interior of P_{n+2} into n triangles. We define the n^{th} Catalan number C_n to be the number of such triangulations of P_{n+2}.
Set $C_0 = 1$. We show in the next section that $C_1 = 1$, $C_2 = 2$ and $C_3 = 5$. Some further values are $C_4 = 14$, $C_5 = 42$, $C_6 = 132$, $C_7 = 429$, $C_8 = 1430$, $C_9 = 4862$ and $C_{10} = 16796$. Six other combinatorial interpretations are also given below.

1.3 Combinatorial Interpretations of Catalan Numbers

We illustrate the combinatorial interpretations assuming $n = 3$ from Stanley's book [11].

1. Figure 1 depicts the triangulations of a convex $(n+2)$-gon into n triangles by $n-1$ diagonals that do not intersect in their interiors.

2. In Figure 2 we look at total number of triangles with vertices 1, i, $i+1$, $2 \leqslant i \leqslant n+1$, among all triangulations of a convex $(n+2)$-gon with vertices $1, 2, ..., n+2$ in clockwise order.

3. A graph is a finite nonempty set V of objects called vertices together with a possible empty set of 2-element subsets of V called edges. A tree is a connected acyclic graph. A vertex v is a child of vertex w if v immediately succeeds w on the

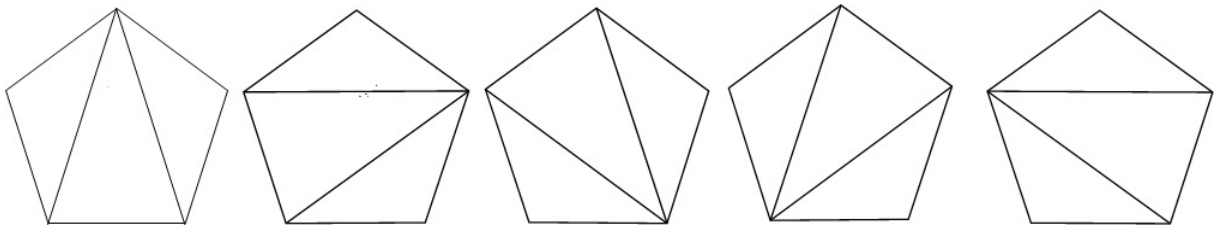

Figure 1: Triangulations of a pentagon by 2 non-intersecting diagonals

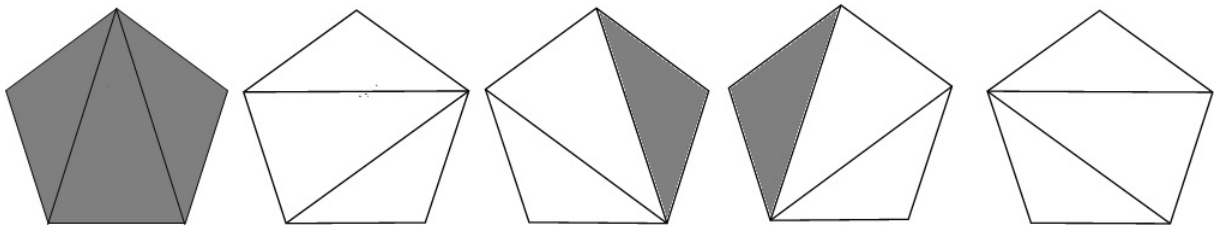

Figure 2: Triangulations of a pentagon into 3 triangles

path from the root to v. See the Graphs & Digraphs book [2] by Chatrand, Lesniak and Zhang for more details. Refer to Figure 3 for the third example about plane trees for which every vertex has 0, 1, or 3 children, with a total of $n+1$ vertices with 0 or 1 child.

4. Lattice path in all generality means a polygonal line of the discrete Cartesian plane $\mathbb{Z} \times \mathbb{Z}$ [1]. In Figure 4, we look at lattice paths from $(0,0)$ to (n,n) with steps $(0,1)$ or $(1,0)$, never rising above the line $y = x$.

5. See Figure 5: Nonnesting matchings on $[2n]$, i.e., ways of connecting $2n$ points in the plane lying on a horizontal line by n arcs, each arc connecting two of the points and lying above the points, such that no arc is contained entirely below another.

6. Sequences a_1, a_2,.....,a_{n-1} of integers such that $a_i \leq 1$ and all partial sums

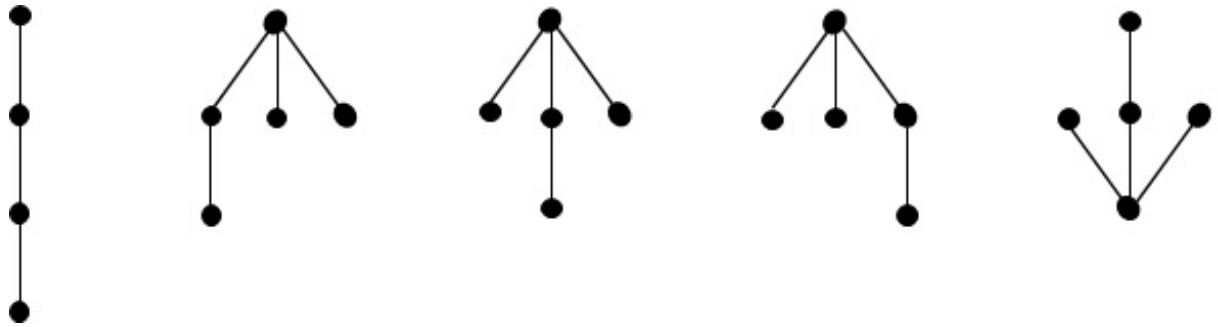

Figure 3: Plane trees with 4 vertices

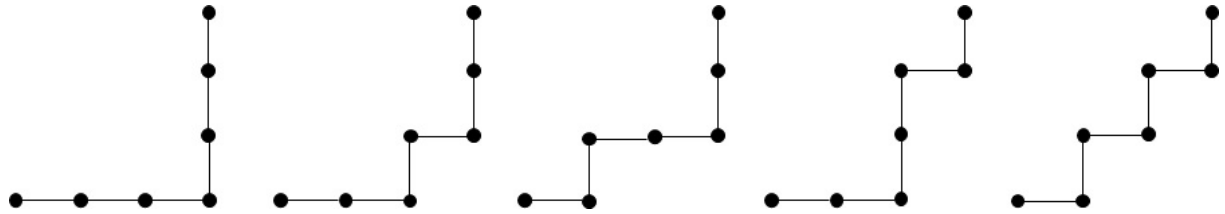

Figure 4: Lattice paths from $(0, 0)$ to $(3, 3)$

are nonnegative are given below:

$$0{,}0 \quad 0{,}1 \quad 1{,}\text{-}1, \quad 1{,}0 \quad 1{,}1$$

In each of the six examples above, we see that $C_3 = 5$. Notice that $C_3 = \frac{1}{3+1}\binom{2\times 3}{3} = 5$ using the formula stated above. In the next section, we will see what this means.

We will also see how the probability mass function of the FPO distribution is related to the so-called Catalan convolutions.

1.4 Catalan Numbers and Catalan Convolutions

In this subsection we first give a formula for the Catalan numbers from the previous subsection, and verify that C_3 is indeed equal to five as calculated in the previous

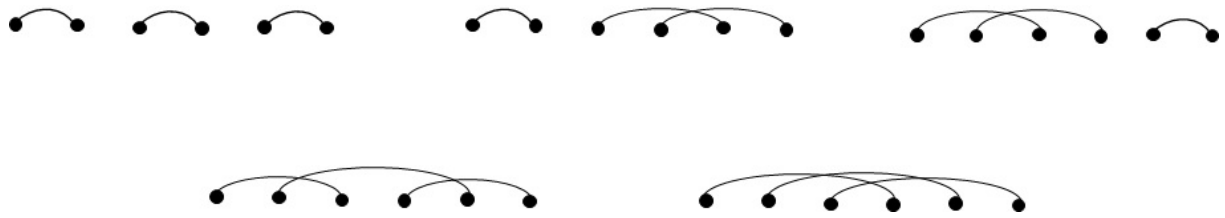

Figure 5: Nonnesting matching on 6 points by 3 arcs

section. From Stanley's book [11], for $n \geq 0$, the Catalan numbers C_n are given by

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

Catalan convolutions are generalization of Catalan numbers.

Example: In Statistics the convolution of two independent random variables is the distribution of their sum. For example, if X and Y are independent discrete non-negative random variables with probability mass function f and g respectively, then the convolution between X and Y can be defined as:

$$(f * g)(n) = \sum f(i)g(n-i)$$
$$\text{or } (f * g)(n) = P(X + Y = n)$$

Generalizing this fact, Catalan proved the k-fold Catalan convolution formula, that is when we have k numbers (instead of two as in the example above) we have:

$$C_{n,k} := \sum_{i_1+\ldots+i_k=n} \prod_{r=1}^{k} C_{i_r-1} = \frac{k}{2n-k}\binom{2n-k}{n} \tag{1}$$

For example, let us look at how to calculate $C_{4,2}$. We proceed as follows: We have three possible values for (i_1, i_2), namely $(1,3)$, $(2,2)$ and $(3,1)$. We thus have $C_{4,2} = C_0 \times C_2 + C_1 \times C_1 + C_2 \times C_0 = 2+1+2 = 5$, which is the same as $C_{4,2} = \frac{2}{(2\times4)-2}\binom{(2\times4)-2}{4}$ as given by the formula.

1.5 Arcsine Distribution with a Lead

Imagine a fair coin toss game between 2 players, Peter and Paul, whose outcome can be modeled by a simple symmetric random walk, graphed in two dimensions with steps to the northeast or southeast. Suppose that the random walk starts at $(m, 0)$ or alternatively Peter starts with a lead of \$m. In other words, we let $Y_i = +/-1$ with probability $\frac{1}{2}$; $Y_0 = m$ and $S_r = \sum_0^r Y_i$; S_r is the position of the random walk after r steps.

Now we are going to flip a coin n times only. We let $N = N_m = min[r; S_r = 0]$, $N = m, m+2, ..., N$ is called the first passage to origin (FPO) distribution. Eventually we would like to study the time X spent by the walk above the t-axis in n steps. This would be the 'arcsine distribution with a lead of m'. When $m = 0$ and $n \to \infty$, this is modeled by the arcsine distribution $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$; $0 < x < 1$. But we are interested in $m > 0$ when the distribution is likely to be left-skewed. This problem was mentioned by a fellow graduate student of the same department, Rebecca Rasnick [8]. There are challenges associated with this problem as we explain below.

First note that the range of N is infinite even though for arcsine distribution we have a finite number of coin flips. How to reconcile this? We proceed as follows:
Note that $X = X_1 + X_2$, where X_1 is the minimum of n and the time N_m taken by the walk to first hit $(0, 0)$, and X_2 is the time spent above the X-axis in the remaining $n - X_1$ units of time. Let us explain this. If for example $n = 20$ and $m = 10$, $N_m = 50$ then Peter has been in lead for all the 20 coin flips; $X_1 = 20$ and $X_2 = 0$. If on the other hand $N_m = 14$, then $X_1 = 14$ and if the remaining tosses are H, H, T, T, T,

H, then $X_2 = 4$ and $X = 18$. We will leave the arcsine distribution with a lead to a future investigation and focus only on the distribution of N_m in this thesis.

1.6 Infinite Expectation

Firstly, $E(N_m) = \infty$ even for $m = 1$; this is not unlike the so-called St. Petersburg paradox [6]; where a player wins 2^m if she takes m flips to get her first head. The distribution of the number of flips is geometric and thus finite with probability one, but the player expects to win ∞. We see the following proof of $E(N_1) \to \infty$.

$$\begin{aligned}
E(N_1) &= \sum_{n \text{ odd}} nP(N = n) \\
&= \sum_{k=1}^{\infty} (2k-1)P(N = 2k-1) \\
&= \sum_{k=1}^{\infty} (2k-1)\frac{1}{2k-1}\binom{2k-1}{k}\frac{1}{2^{2k-1}} \\
&\geq \frac{C * 2^{2k}}{\sqrt{k} * 2^{2k-1}} \\
&= \sum \frac{C}{\sqrt{k}} \\
&\to \infty
\end{aligned}$$

where the last inequality is proved using the Stirling's approximation for factorials. That is, $k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$. In fact we have, $\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \leq k! \leq \sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\frac{1}{12k}}$. [10]

In fact as we see by Feller, W. in [5], the probability mass function of the FPO distribution is given by:

$$P(N_m = n) = \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n};\ n \equiv m \pmod 2. \tag{2}$$

With $m = 4$, for example, we get for $k \geq 2$ (by (2))

$$P(N_4 = 2k) = \frac{2}{k}\binom{2k}{k+2}\frac{1}{2^{2k}},$$

which simplifies on setting $k = r - 2$ for $r \geq 4$ as

$$P(N_4 = 2r-4) = \frac{4}{2r-4}\binom{2r-4}{r}\frac{1}{2^{2r-4}} = \frac{C(r,4)}{2^{2r-4}},$$

and we see the 4th Catalan convolution of r appears. Remember the formula for $C_{n,k} = \frac{k}{2n-k}\binom{2n-k}{n}$ from (1) above. It turns out that for general m we get the mth Catalan convolution!

For odd m, $m = 2s + 1$, we see that we must have $n = 2k + 1$, $k \geq s$ [3]. This is true since if initial lead $\$m$ is odd, then the number of steps to go bankrupt, n, shall be odd too. Thus by (2)

$$P(N_{2s+1} = 2k+1) = \frac{2s+1}{2k+1}\binom{2k+1}{k+s+1}\frac{1}{2^{2k+1}},$$

which simplifies on setting $k = r - s$ with $r \geq 2s$, as

$$\begin{aligned} P(N_{2s+1} = 2(r+1)-(2s+1)) &= \frac{2s+1}{2(r+1)-(2s+1)}\binom{2(r+1)-(2s+1)}{r+1}\frac{1}{2^{2(r+1)-(2s+1)}} \\ &= \frac{C(r+1,2s+1)}{2^{2(r+1)-(2s+1)}} \end{aligned}$$

Similarly, for even m, $m = 2s + 2$, we must have $n = 2k + 2$, $k \geq s$. Thus

$$P(N_{2s+2} = 2k+2) = \frac{2s+2}{2k+2}\binom{2k+2}{k+s+1}\frac{1}{2^{2k+2}},$$

which simplifies on setting $k = r - s$ with $r \geq 2s$, as

$$\begin{aligned} P(N_{2s+2}) &= 2(r+2)-(2s+2)) \\ &= \frac{2s+2}{2(r+2)-(2s+2)}\binom{2(r+2)-(2s+2)}{r+1}\frac{1}{2^{2(r+2)-(2s+2)}} \\ &= \frac{C(r+2,2s+2)}{2^{2(r+2)-(2s+2)}} \end{aligned}$$

To conclude, we see that the FPO distribution does involve the Catalan Convolutions. Next, we shall see the relation between the FPO distribution and the Ballot Theorem.

1.7 FPO Distribution and Ballot Theorem

The Ballot Theorem: Let n and x be positive numbers. There are $\frac{x}{n}N_{n,x}$ number of paths that are strictly above the t-axis for $t > 0$ that join $(0, 0)$ to (n, x) where $N_{n,x} = \binom{n}{\frac{n+x}{2}}$.

FPO distribution can be seen as the Ballot theorem being looked backwards. In FPO distribution, we are looking at time to reach origin when we already have a lead of $\$m$, i.e. we start at $(m, 0)$. See figure 6.

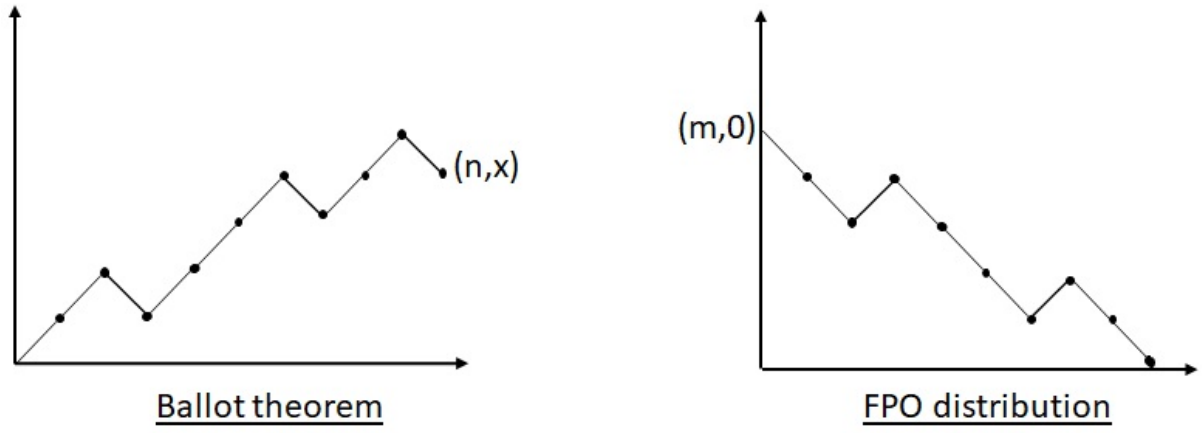

Figure 6: FPO distribution and Ballot theorem .

For this distribution, we flip things around and consider $(n + x)$ as a random variable now. $(n+x)$ is essentially the time to return to origin. In other words, P(the paths are strictly above the t-axis and join $(0, 0)$ to $(n, x)) = \frac{x}{n}\binom{n}{\frac{n+x}{2}}\frac{1}{2^n}$.

We thus have, P(the paths are strictly above the t-axis and join $(0, m)$ to $(n, 0)$) $= \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n}$.

Next, we look at the proof of the Catalan Convolution formula as in (1) above.

1.8 Proof of Catalan Convolution Formula

In this section, we prove the Catalan Convolution formula explaining the method used by Regev, A. in [9]. The author proceeds by a sequence of lemmas. We provide complete details here.

Lemma 1 A k-dissection of an n-gon is a partition of the n-gon into k + 1 parts by k noncrossing diagonals. The number of $k - in - n$ dissections is:

$$f_k(n) = \binom{2n-k-1}{n-1} \tag{3}$$

where a $k - in - n$ dissections is

Lemma 2 Let $3 \leq k \leq n$, Then

$$(n-k)f_k(n) = n\sum_{i=2}^{n-k+1} C_{i-1}f_k(n-i+1). \tag{4}$$

Lemma 3 For any $n \geq 1$,

$$\sum_{i\geq 0} iC_iC_{n-i} = \binom{2n+1}{n-1}. \tag{5}$$

Lemma 4 Let $1 \leq q \leq p \leq 2q-1$. Then

$$\sum_{i\geq 0} C_i\binom{p-1-2i}{q-1-i} = \binom{p}{q}. \tag{6}$$

Lemma 5 Let $3 \leq k \leq n$. Then

$$kf_k(n) = n \sum_{i_1+....+i_k=n} C_{i_1-1}....C_{i_k-1}. \tag{7}$$

The lemmas yield the following theorem:

Theorem 1.1 *Let $1 \leq k \leq n$. Then*

$$\sum_{i_1+....+i_k=n} C_{i-1}....C_{i_k-1} = \frac{k}{2n-k}\binom{2n-k}{n}.$$

Proof: Fix $k \geq 3$ and proceed by induction on n. If $n = k$ then both sides are equal to 1. Now let $n \geq k+1$. From Lemma 1 and Lemma 2 and using the induction hypothesis, we have

$$\begin{aligned}
f_k(n) &= \frac{n}{n-k}\sum_{i=2}^{n-k-1} C_{i-1}f_k(n-i+1) \\
&= \frac{n}{n-k}\sum_{i=2}^{n-k-1} C_{i-1}\binom{2(n-i+1)-k-1}{n-i} \\
&= \frac{n}{n-k}\left(\sum_{i=1}^{n-k-1} C_{i-1}\binom{2(n-i+1)-k-1}{n-i} - C_{i-1}\binom{2(n-1+1)-k-1}{n-1}\right) \\
&= \frac{n}{n-k}\left(\sum_{i\geq 1} C_{i-1}\binom{2(n-i+1)-k-1}{n-i} - f_k(n)\right)
\end{aligned}$$

Solving for $f_k(n)$, we get

$$f_k(n)\left(\frac{2n-k}{n-k}\right) = \frac{n}{n-k}\left(\sum_{i\geq 1} C_{i-1}\binom{2(n-i+1)-k-1}{n-i}\right)$$

$$f_k(n) = \frac{n}{2n-k}\sum_{i\geq 0} C_i\binom{2n-k-2i-1}{n-i-1}$$

Using Lemma 4 with $q = n$ and $p = 2n - k$, we see

$$f_k(n) = \frac{n}{2n-k}\binom{2n-k}{n}$$
$$f_k(n) = \binom{2n-k-1}{n-1}$$

Now using Lemma 5, we get

$$\sum_{i_1+....+i_k=n} C_{i_1-1}....C_{i_k-1} = \frac{k}{n}\binom{2n-k-1}{n-1}$$
$$= \frac{k}{2n-k}\binom{2n-k}{n}.$$

This proves the Theorem 1.1.

Let us move on to some more properties of the FPO distribution.

2 Exploratory Analysis of the FPO Distribution

In Chapter 1 we saw,

$$P(N_m = n) = \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n}; n \equiv m \pmod 2$$

There are some cognate results to look at:

$$\sum_n P(N = n) = 1 \tag{8}$$

We argue that the above relation is true as eventually the random walk should hit the origin with a probability of 1. This can be proved using Polya's theorem [7] on recurrence of random walks. We skip the proof in this thesis but can be accessed from [7].

To get some numerical evidence for equation (8), we used R to plot histogram for the above pmf with $m = 1$ and odd values of n. We get the histogram as in Figure 7 on the next page.

It can be shown that the cumulative probability when n varies from $n = 1$ to $n = 515$, sums to 0.975. This also gives us evidence that $\sum P(N = n) = 1$.

The quantiles for the FPO distribution will be as follows: median value is at $n = 1$; 75th percentile will be at $n = 5$; and 90th percentile will be at $n = 32$.

In the next section, we look at the distribution of N.

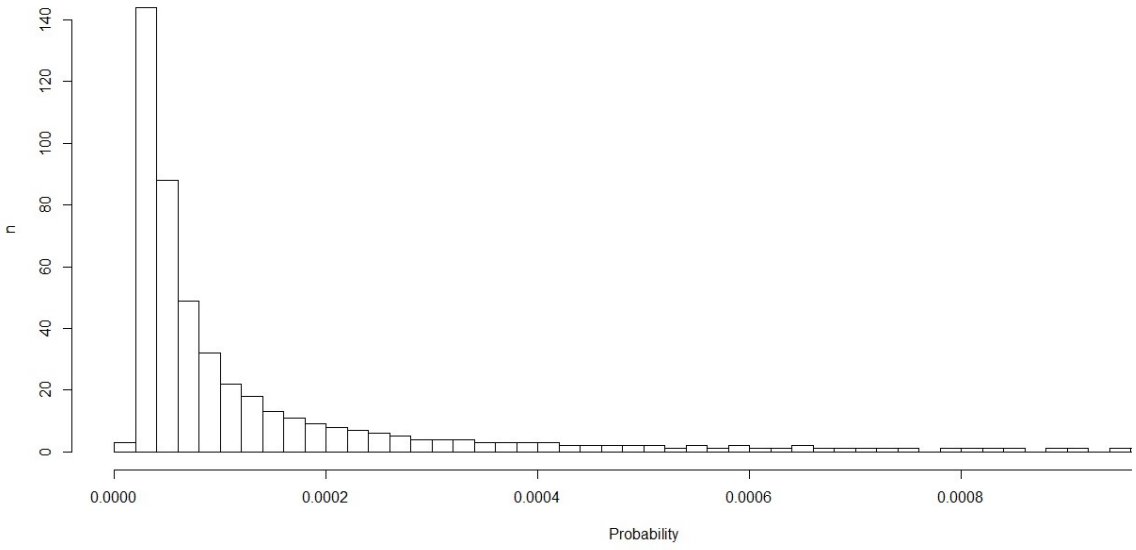

Figure 7: Histogram of the FPO distribution .

3 Distribution of $N = N_m$

We seek to approximate $P(N = n)$ under various conditions for large values of m and n. First assume, $\frac{n+m}{2}$ is close to n. This occurs if,

$$\frac{m}{n} \sim 1, \text{ or if}$$

$$\frac{m}{n} = 1 - \epsilon(n) \text{ for some } \epsilon(n) > 0 \text{ , or if}$$

$$\frac{n}{2} + \frac{m}{2} = n - \phi(n) \text{ where } \frac{\phi(n)}{n} \to 0$$

Under this condition, we now look at the exact and approximate distribution of N.

3.1 Exact and Approximate Distributions of N

The exact distribution now becomes,

$$\begin{aligned} P(N=n) &= \frac{n-n\epsilon(n)}{n}\binom{n}{n-\phi(n)}\frac{1}{2^n} \\ &= \frac{n(1-\epsilon(n))}{n}\binom{n}{\phi(n)}\frac{1}{2^n} \end{aligned}$$

Using Stirling's approximation and assuming $\phi(n) \ll n$, the approximate distribution will be:

$$\begin{aligned} P(N=n) &\simeq \frac{n^{\phi(n)}}{\phi(n)!}\frac{1}{2^n}, \text{ or} \\ P(N=n) &\simeq \frac{n^{\phi(n)}}{\sqrt{2\pi\phi(n)}}(\frac{e}{\phi(n)})^{\phi(n)} \end{aligned} \tag{9}$$

If $\phi(n)$ is really big, that is $\phi(n) = \frac{n}{2}$ - $\frac{m}{2} = Bn$ for some large B, then we cannot use the approximation $\binom{n}{\phi(n)} \sim \frac{n^{\phi(n)}}{\phi(n)!}$. This approximation is valid only if $\phi(n)$ is big but not as large as Bn.

Let us consider certain other cases when $\phi(n) \neq Bn$ and look at the exact and approximate distributions for each case, so as to understand how good the approximation is.

CASE 1: In the first case and easiest case, we let $n = m$.

In this case, $\epsilon(n) = \phi(n) = 0$

and, the exact distribution = approximate distribution $\sim \frac{1}{2^m}$.

CASE 2: $n = m + 2$. Here $\epsilon(n) = \frac{2}{n}$, $\phi(n) = 1$.

The exact distribution will be:

$$\begin{aligned} P(N=n) &= \frac{m}{m+2}\binom{m+2}{m+1}\frac{1}{2^{m+2}} \\ &= \frac{m}{m+2}\frac{m+2}{2^{m+2}} \\ &= \frac{m}{2^{m+2}}. \end{aligned}$$

and the approximate distribution is

$$\begin{aligned} P(N=n) &= \frac{m+2}{1!}\frac{1}{2^{m+2}} \\ &\sim \frac{m}{2^{m+2}}. \end{aligned}$$

CASE 3: $n = m + 50$

The exact distribution is

$$\begin{aligned} P(N=n) &= \frac{m}{m+50}\binom{m+50}{\frac{2m+50}{2}}\frac{1}{2^{m+50}} \\ &= \frac{m}{m+50}\binom{m+50}{25}\frac{1}{2^{m+50}}. \end{aligned}$$

and the approximate distribution is $\sim \frac{(m+50)^{25}}{25!}\frac{1}{2^{m+50}}$.

Next, we proceed to the probability mass function of N.

3.2 Probability Mass Function of N

We first calculate the probability of N lying between am^2 and bm^2 where a and b are positive constants, and show that this tends to 1 as $a \to 0$ and $b \to \infty$. We shall assume that $\frac{m}{n}$ is small for reasons that we clarify later.

We first simplify the pmf for N. That is,

$$\begin{aligned}
P(N=n) &= \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n} \\
&= \frac{m}{n}\frac{1}{2^n}\frac{n!}{(\frac{n+m}{2})!(\frac{n-m}{2})!} \\
&\simeq \frac{m}{n}\frac{1}{2^n}\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(\frac{2e}{n+m}\right)^{\frac{n+m}{2}}\left(\frac{2e}{n-m}\right)^{\frac{n-m}{2}}\frac{1}{\sqrt{\frac{2\pi(n+m)}{2}}}\frac{1}{\sqrt{\frac{2\pi(n-m)}{2}}} \\
&= \frac{m}{n}\sqrt{\frac{2n}{\pi}}\frac{1}{\left(1+\frac{m}{n}\right)^{\frac{n+m}{2}}\left(1-\frac{m}{n}\right)^{\frac{n-m}{2}}}\frac{1}{\sqrt{(n+m)(n-m)}} \\
&\sim \frac{m}{n^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\frac{1}{\left(1+\frac{m}{n}\right)^{\frac{n+m}{2}}\left(1-\frac{m}{n}\right)^{\frac{n-m}{2}}}
\end{aligned}$$

We use the approximations $e^{xy} \simeq (1+x)^y$ and $\sqrt{(n+m)(n-m)} \simeq \sqrt{n^2}$. These approximations are valid because $x = \frac{m}{n}$ is small. Using these approximations, we get:

$$\begin{aligned}
P(N=n) &\sim \frac{m}{n^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\frac{1}{e^{\frac{m}{n}(\frac{n+m}{2})}e^{\frac{-m}{n}(\frac{n-m}{2})}} \\
&= e^{\frac{-m^2}{n}}\frac{m}{n^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}} \qquad (10) \\
\text{So that } P(am^2 \le n \le bm^2) &= \sum_{am^2}^{bm^2}\sqrt{\frac{2}{\pi}}\frac{m}{n^{\frac{3}{2}}}e^{\frac{-m^2}{n}}
\end{aligned}$$

We assume that n is even, so replacing n with $2n$ and considering only even values for n, we see that,

$$\begin{aligned}
P(am^2 \le N \le bm^2) &= \sum_{am^2/2}^{bm^2/2}\sqrt{\frac{2}{\pi}}\frac{m}{2^{\frac{3}{2}}n^{\frac{3}{2}}}e^{\frac{-m^2}{2n}} \\
&\simeq \sqrt{\frac{2}{\pi}}\int_{am^2/2}^{bm^2/2}\frac{m}{2^{\frac{3}{2}}n^{\frac{3}{2}}}e^{\frac{-m^2}{2n}}dn
\end{aligned}$$

We solve this by substitution: Let $u^2 = \frac{m^2}{n}$, $u = \frac{m}{\sqrt{n}}$, $du = m(\frac{-1}{2})n^{\frac{-3}{2}}dn$

The integral now becomes:

$$P(am^2 \leq N \leq bm^2) = -\sqrt{\frac{2}{\pi}}\frac{1}{2^{\frac{3}{2}}}\int_{\sqrt{\frac{2}{a}}}^{\sqrt{\frac{2}{b}}} 2e^{\frac{-u^2}{2}}du$$
$$= \sqrt{\frac{2}{\pi}}\frac{2}{2^{\frac{3}{2}}}\int_{\sqrt{\frac{2}{b}}}^{\sqrt{\frac{2}{a}}} e^{\frac{-u^2}{2}}du$$

Assuming $a \to 0$ and $b \to \infty$, the above integral tends to

$$= \sqrt{2}\int_0^{\infty} \frac{1}{\sqrt{2\pi}}e^{\frac{-u^2}{2}}du$$
$$= \sqrt{2} * \frac{1}{2}$$
$$\simeq 0.707$$

which we will show is a result of error in approximating the series by integral. That is, we claim and show later that $P(am^2 \leq N \leq bm^2) \to 1$ as $a \to \infty$ and $b \to \infty$.

Next, let us find the probability of n lying between am and ∞. We have,

$$P(am \leq N < \infty) \simeq \sqrt{\frac{2}{\pi}}\int_{am}^{\infty} \frac{m}{2^{\frac{3}{2}}n^{\frac{3}{2}}}e^{\frac{-m^2}{2n}}dn$$

Substituting $u^2 = \frac{m^2}{n}$, the above integral becomes

$$-\sqrt{\frac{2}{\pi}}\int_{\sqrt{\frac{m}{a}}}^{0} e^{\frac{-u^2}{2}}\frac{1}{\sqrt{2}}du \text{ , or}$$
$$\sqrt{2}\int_0^{\sqrt{\frac{m}{a}}} e^{\frac{-u^2}{2}}\frac{1}{\sqrt{2\pi}}du \text{ , or}$$

When a $\to 0$, we see $P(am \leq N \leq \infty)$ tends to

$$\sqrt{2} * \frac{1}{2} \simeq 0.707$$

which is same as in the previous calculation and we see that no gain is made by extending the range of N to am.

Now, let us assume n is a function of m, say $n = \phi(m)m$, where $\phi(m) \to \infty$ as $m \to \infty$. Hence,

$$P(\phi(m)m \leq N < \infty) = \sum_{\phi(m)m}^{\infty} \sqrt{\frac{2}{\pi}} \frac{m}{n^{\frac{3}{2}}} e^{\frac{-m^2}{n}}$$

Replace n with $2n$ as we are only looking at even values for n.

$$\begin{aligned} P(\phi(m)m \leq N < \infty) &= \sum_{\phi(m)m/2}^{\infty} \sqrt{\frac{2}{\pi}} \frac{m}{2^{\frac{3}{2}} n^{\frac{3}{2}}} e^{\frac{-m^2}{2n}} \\ &\simeq \sqrt{\frac{2}{\pi}} \int_{\frac{m\phi(m)}{2}}^{\infty} \frac{m}{2^{\frac{3}{2}} n^{\frac{3}{2}}} e^{\frac{-m^2}{2n}} dn \end{aligned}$$

Solving this by substitution, as before, the integral now becomes:

$$\begin{aligned} &= -\sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2}} \int_{\sqrt{\frac{2m}{\phi(m)}}}^{0} e^{\frac{-u^2}{2}} du \\ &= \sqrt{\frac{2}{\pi}} \sqrt{2} \int_{0}^{\sqrt{\frac{2m}{\phi(m)}}} e^{\frac{-u^2}{2}} du \\ &= \sqrt{2} * \frac{1}{2} \\ &\simeq 0.707 \end{aligned}$$

as $\phi(m) \to \infty$ which is again the same as before.

We observed that the probability that we got in each of the three cases above is ~ 0.707. We claimed that this value is different from 1 due to error by approximating a series with an integral. We prove this claim by looking at the probability of any interval outside the three intervals considered above and showing that the probability of each of those interval $\to 0$.

We now present our claims:

Claim 1: P$(\phi(m)m \leq N \leq am^2) \to 0$ as $a \to 0$ and $m \to \infty$

This is true because $P(am^2 \leq N \leq bm^2)$, when approximated by an integral, tends

to 0.707 as $a \to 0$ and $b \to \infty$. And $P(m\phi(m) \leq N < \infty)$ can also be approximated by 0.707. Thus,

$$P(m\phi(m) \leq N \leq am^2) = P(m\phi(m) \leq N < \infty) - P(am^2 \leq N < \infty) \to 0$$

Claim 2: $P(N \geq bm^2) \to 0$ as $b \to \infty$

Proof:

$$\begin{aligned} P(N \geq bm^2) &\simeq \sqrt{\frac{2}{\pi}} \int_{\frac{bm^2}{2}}^{\infty} \frac{m}{2^{\frac{3}{2}} n^{\frac{3}{2}}} e^{\frac{-m^2}{2n}} dn \\ &= -\sqrt{\frac{2}{\pi}} \sqrt{\frac{1}{2}} \int_{\sqrt{\frac{2}{B}}}^{0} e^{\frac{-u^2}{2}} du \\ &\to 0 \text{ as } b \to \infty \end{aligned}$$

Next, let us consider the case $n = Bm$ using the exact distribution,

$$P(N = n) = \frac{1}{2^n} \frac{m}{n} \binom{n}{\frac{m+n}{2}}$$

In this case, we cannot use the approximation $(1+x) \sim e^x$ for small x, or $\sqrt{(n+m) * (n-m)} \sim n$. Letting $m = \frac{n}{B}$, we have

$$\begin{aligned} P(N = n) &= \frac{1}{2^n} \frac{m}{mB} \binom{n}{\frac{n+\frac{n}{B}}{2}} \\ &= \frac{1}{2^n} \frac{1}{B} \frac{n!}{\left(n\frac{B+1}{2B}\right)! \left(n\frac{B-1}{2B}\right)!} \\ &= \frac{1}{2^n} \frac{1}{B} \frac{\sqrt{2n\pi} \frac{n^n}{e^n} e^{\frac{n(B+1)}{2B}} e^{\frac{n(B-1)}{2B}}}{\sqrt{2n\pi \frac{B+1}{2B}} \sqrt{2n\pi \frac{B-1}{2B}} \left(\frac{n(B+1)}{2B}\right)^{\frac{n(B+1)}{2B}} \left(\frac{n(B-1)}{2B}\right)^{\frac{n(B-1)}{2B}}} \\ &= \frac{2B}{B} \frac{1}{\sqrt{2n\pi}} \frac{B^n}{\sqrt{B^2-1}} \frac{1}{(B+1)^{\frac{n(B+1)}{2B}} (B-1)^{\frac{n(B-1)}{2B}}} \\ &= \frac{1}{B} \frac{2}{\sqrt{2mB\pi}} \left(\left(\frac{B}{B+1}\right)^{\frac{B+1}{2}} \left(\frac{B}{B-1}\right)^{\frac{B-1}{2}} \right)^m \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{m}} \frac{1}{B^{\frac{3}{2}}} \left(\left(\frac{1}{1+\frac{1}{B}}\right)^{\frac{B+1}{2}} \left(\frac{1}{1-\frac{1}{B}}\right)^{\frac{B-1}{2}} \right)^m \end{aligned}$$

Since, $1 + \frac{1}{B} \leq e^{\frac{1}{B}}$, we have $\frac{1}{(1+\frac{1}{B})} \geq \frac{1}{e^{1/B}}$ and $\frac{1}{(1-\frac{1}{B})} \geq \frac{1}{e^{-1/B}}$, so that

$$\begin{aligned} P(N=n) &\geq \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{m}} \frac{1}{B^{3/2}} \left(\left(e^{1/B}\right)^{\frac{B+1}{2}} \left(e^{-1/B}\right)^{\frac{B-1}{2}} \right)^{-m} \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{m}} \frac{1}{B^{3/2}} \left(e^{\frac{B+1-B+1}{2B}} \right)^{-m} \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{m}} \frac{1}{B^{\frac{3}{2}}} e^{\frac{-m}{B}} \end{aligned}$$

Thus, approximating $P(m \leq N \leq am)$ using integration:

$$P(m \leq N \leq am) \geq \sqrt{\frac{2}{\pi}} \int_1^a \frac{1}{\sqrt{m}} \frac{1}{B^{3/2}} e^{\frac{-m}{B}} dB$$

Let $\frac{m}{B} = \frac{u^2}{2}$, $u = \sqrt{\frac{2m}{B}}$ and $du = \sqrt{2m} \frac{-1}{2} B^{\frac{-3}{2}} dB$, so

$$\begin{aligned} P(m \leq N \leq am) &= -2\sqrt{\frac{2}{\pi}} \int_{\sqrt{2m}}^{\sqrt{2m/a}} \frac{1}{\sqrt{m}} \frac{B^{3/2}}{B^{3/2}} e^{\frac{-u^2}{2}} du \\ &= \frac{2}{\sqrt{2}} \sqrt{\frac{2}{\pi}} \int_{\sqrt{2m/a}}^{\sqrt{2m}} e^{\frac{-u^2}{2}} du \\ &\geq \frac{2}{\sqrt{2}} \sqrt{\frac{2}{\pi}} \sqrt{2m} \left(1 - \frac{1}{a}\right) e^{-m} \end{aligned}$$

which is small as $m \to \infty$. A similar small upper bound for $P(m \leq N \leq am)$ is obtained by using the inequality $(1+u) \geq e^{\frac{u}{(1+u)}}$.

Claim 3: $P(N \leq am) \to 0$ as $a \to \infty$

Using same substitution as before:

$$\begin{aligned} P(N \leq am) &\simeq \sqrt{\frac{2}{\pi}} \int_0^{am} \frac{m}{2^{\frac{3}{2}} n^{\frac{3}{2}}} e^{\frac{-m^2}{2n}} dn \\ &= -\sqrt{\frac{2}{\pi}} \sqrt{\frac{1}{2}} \int_{\infty}^{\sqrt{am}} e^{\frac{-u^2}{2}} du \\ &= \sqrt{\frac{2}{\pi}} \sqrt{\frac{1}{2}} \int_{\sqrt{am}}^{\infty} e^{\frac{-u^2}{2}} du \\ &\to 0 \text{ as } a \to \infty \end{aligned}$$

Claims 1, 2 and 3 lead to our main result:

Theorem 3.1 *$P(am^2 \leq N \leq bm^2) \to 1$ as $a \to 0$, $b \to \infty$ and $m \to \infty$. In other words, the distribution of N is concentrated in an interval of length $\Theta(m^2)$.*

Next, let us look at the mode of N, assuming that m is known.

Theorem 3.2 *The mode of N (given m) is $\frac{m^2-4}{3}$.*

Proof: We set

$$\phi(n) = \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n} \ ,$$

and use the fact that $\phi(n) = P(N = n)$ will be increasing when the ratio $\frac{\phi(n+2)}{\phi(n)} \geq 1$. So, simplifying the ratio as follows

$$\begin{aligned}\frac{\phi(n+2)}{\phi(n)} &= \frac{n}{n+2}\frac{\binom{n+2}{\frac{n+m+2}{2}}}{\binom{n}{\frac{n+m}{2}}}\frac{1}{4} \\ &= \frac{n}{4n+8}\frac{(n+2)(n+1)}{\frac{n+m+2}{2}}\frac{\frac{(n-m)!}{2!}}{\frac{(n-m+2)!}{2!}}\end{aligned}$$

we set,

$$\frac{n}{4n+8}\frac{(n+2)(n+1)}{\frac{n+m+2}{2}}\frac{\frac{(n-m)!}{2!}}{\frac{(n-m+2)!}{2!}} \geq 1$$

to find the maximum of the pmf.

$$\begin{aligned}\text{That is, } \frac{n(n+1)}{4}\frac{1}{(\frac{n+m+2}{2})(\frac{n-m+2}{2})} &\geq 1 \text{ iff} \\ n^2 + n &\geq (n+2)^2 - m^2 \text{ iff} \\ 3n - m^2 + 4 &\leq 0 \text{ iff} \\ n &\leq \frac{m^2-4}{3}\end{aligned}$$

Hence, the mode of N (given m) is $\frac{m^2-4}{3}$.

We now study the challenges faced while computing the quantiles of N.

3.3 Why are Quantiles of N Difficult to Calculate?

With large m, we computed an approximate distribution in an effort to understand N better. However, in order to find different probabilities, we approximated the sum with an equivalent integral. Thus, the actual probabilities, that is the sum of the rectangles in the picture below, will always be more than the area under the curve. Hence, by approximating the distribution with an integral, we will see that the approximated probability will always be less than actual probability. Remember, from section 3.2 above, we got the total probability of about 0.707. However it should have been 1 as seen by Theorem 3.1. Additionally, it is very difficult to handle this

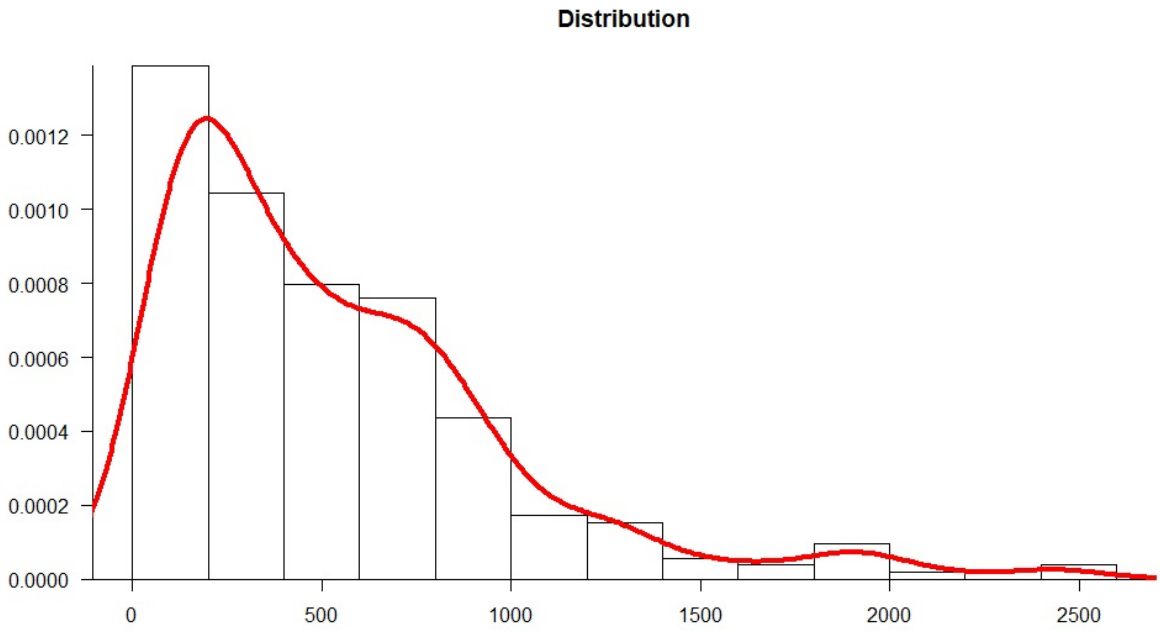

Figure 8: Approximating series with an integral

distribution computationally as it involves large factorials and combinations. We had used software (R) to do this for us. We could only get results for n from 1 to 515 and even that only for $m = 1$. In Chapter 1 we found certain percentiles using results from the software.

In conclusion,

1. Whenever we do an integral approximation to discrete case, the approximate probability will always be lower than the actual probability.

2. It will be really difficult computationally to handle factorials for large m. Thus, we had used software to handle this for us but only for small m.

In the last section we calculate the maximum likelihood estimate for m.

4 Maximum Likelihood Estimation of m

In this section we assume that m is unknown and attempt to find the MLE of m given k sample values n_1, n_2,....,n_k, that is, k first passage to the origin samples.

k=1: First, let us find the Maximum Likelihood Estimate for m when $k = 1$. The likelihood function in this case is:

$$L = \frac{m}{n} \frac{\binom{n}{\frac{n+m}{2}}}{2^n} = \phi(m)$$

We maximize L with respect to the parameter m and find the corresponding MLE for m. So we see when ϕ is increasing by asking when $\frac{\phi(m+2)}{\phi(m)}$ is greater than one.

$$\frac{\phi(m+2)}{\phi(m)} = \frac{m+2}{m} \frac{\binom{n}{\frac{n+m+2}{2}}}{\binom{n}{\frac{n+m+2}{2}}} \geq 1 \text{ iff}$$

$$\frac{m+2}{m} \frac{\frac{n-m}{2}}{\frac{n+m+2}{2}} \geq 1 \text{ iff}$$

$$nm - m^2 + 2n - 2m \geq nm + m^2 + 2m \text{ iff}$$

$$m^2 + 2m - n \leq 0$$

The roots of the corresponding quadratic equation are $-\sqrt{n+1}-1 \leq m \leq \sqrt{n+1}-1$. Thus the quadratic is negative between the two roots. Since m can only take positive values, the value of m at which L is maximized is between 0 and -1 + $\sqrt{1+n}$, leading to the conclusion that MLE $\hat{m}$ is around $\sqrt{n}$.

k=2: Now, let us find the Maximum Likelihood Estimate for m when $k = 2$. The likelihood function $L = \phi(m)$ is

$$\phi(m) = \frac{m^2}{n_1 n_2} \frac{\binom{n_1}{\frac{n_1+m}{2}} \binom{n_2}{\frac{n_2+m}{2}}}{2^{n_1 n_2}}$$

Thus,

$$\frac{\phi(m+2)}{\phi(m)} = \frac{(m+2)^2}{n_1 n_2} \frac{\binom{n_1}{\frac{n_1+m+2}{2}}\binom{n_2}{\frac{n_2+m+2}{2}}}{2^{n_1 n_2}} \Big/ \frac{m^2}{n_1 n_2} \frac{\binom{n_1}{\frac{n_1+m}{2}}\binom{n_2}{\frac{n_2+m}{2}}}{2^{n_1 n_2}}$$

$$= \frac{\binom{m+2}{m}^2 \frac{n_1-m}{2} \frac{n_2-m}{2}}{\frac{n_1+m+2}{2} \frac{n_2+m+2}{2}}$$

$$= \frac{(m^2+4m+4)(n_1 n_2 - n_1 m - n_2 m + m^2)}{m^2(n_1 n_2 + n_1 m + 2n_1 + m n_2 + m^2 + 2m + 2n_2 + 2m + 4)}$$

Thus, $\frac{\phi(m+2)}{\phi(m)} \geq 1$ iff

$$\frac{(m^2+4m+4)(n_1 n_2 - n_1 m - n_2 m + m^2)}{m^2(n_1 n_2 + n_1 m + 2n_1 + m n_2 + m^2 + 2m + 2n_2 + 2m + 4)} \geq 1 \text{ iff}$$

$$m^2 n_1 n_2 - n_1 m^3 - m^3 n_2 + m^4 + 4m n_1 n_2 - 4n_1 m^2 - 4m^2 n_2 + 4m^3 + 4n_1 n_2 - 4n_1 m$$

$$-4m n_2 + 4m^2 \geq m^2 n_1 n_2 + m^3 n_1 + 2n_1 m^2 + m^3 n_2 + m^4 + 2m^3 + 2n_2 m^2 + 2m^3 + 4m^2 \text{ iff}$$

$$m^3(n_1+n_2) + 3m^2(n_1+n_2) - 2n_1 n_2 - 2m n_1 n_2 + 2(n_1+n_2)m \leq 0$$

Solving the above cubic equation equation will give us the maximum likelihood estimate for m when $k = 2$. We restrict to the case where $n_1 = n_2$ since the general case is quite complex.

Suppose, $n_1 = n_2 =$ n. The above condition will now be:

$$2m^3 n + 6m^2 n - 2n^2 - 2mn^2 + 4nm \leq 0 \text{ iff}$$

$$m^3 n + 3m^2 n - n^2 - mn^2 + 2nm \leq 0 \text{ iff} \tag{11}$$

$$m^3 + 3m^2 - n - mn + 2m \leq 0$$

We couldn't solve this analytically. Using software, the Maximum Likelihood Estimate for m or in other words the largest m for which (11) holds so we look for the largest m such that:

$$-1 \leq m \leq \sqrt{n+1} - 1 \ ,$$

or, again

$$\hat{m} \sim \sqrt{n}$$

In an effort to understand the general case, let us estimate the MLE for m for some random values of n_1 and n_2: (See Table 1)

Table 1: MLE for m when $k = 2$ for some random values of n_1 and n_2

n1	**n2**	**m**
10	20	3
20	20	3
200	10	2
50	300	4
100	500	11
40	1000	4
1000	10000	41

Let us also look at some of the sample values for $n_1 = n_2 = n$ and the corresponding maximum likelihood estimates for m: (See Table 2)

Table 2: MLE for m when $k = 2$ assuming $n_1 = n_2 = n$

n	**Sqrt(n)**	**m**
10	3.16	2
20	4.47	3
100	10	9
250	15.81	14
500	22.36	21
1000	31.63	30
10000	100	99
340	18.44	17
269	16.40	15
1056	32.50	31
1200	34.64	33
10000000	3162.28	3161

We can see from the above table, when $n_1 = n_2 = n$, then the maximum likelihood

value of m is almost equal to the square root of n but that the general case is far more advanced.

k=3: Now, let us find the Maximum Likelihood Estimate for m when $k = 3$ and $n_1 = n_2 = n_3 = n$.

$$\phi(m) = \frac{m^3}{n^3}\binom{n}{\frac{n+m}{2}}^3 \frac{1}{8^n}$$

$$\text{So, } \frac{\phi(m+2)}{\phi(m)} = \frac{\frac{(m+2)^3}{n^3}\binom{n}{\frac{n+m+2}{2}}^3 \frac{1}{8^n}}{\frac{m^3}{n^3}\binom{n}{\frac{n+m}{2}}^3 \frac{1}{8^n}}$$

$$= \frac{(m+2)^3}{m^3}\left(\frac{(\frac{n+m}{2})!(\frac{n-m}{2})!}{(\frac{n+m+2}{2})!(\frac{n-m-2}{2})!}\right)^3$$

Thus, $\frac{\phi(m+2)}{\phi(m)} \geq 1$ if and only if

$$\frac{(m+2)^3}{m^3}\left(\frac{(\frac{n+m}{2})!(\frac{n-m}{2})!}{(\frac{n+m+2}{2})!(\frac{n-m-2}{2})!}\right)^3 \geq 1 \text{ , or}$$

$$\frac{(m+2)^3}{m^3} \geq \left(\frac{(\frac{n+m+2}{2})!(\frac{n-m-2}{2})!}{(\frac{n+m}{2})!(\frac{n-m}{2})!}\right)^3$$

Solving the above inequality, we again see that the maximum likelihood estimate for m is about square root of n for $k = 3$.

In Table 3, we estimate the MLE for m for some values when $n_1 = n_2 = n_3 = n$.

k=4: Now, let us find the Maximum Likelihood Estimate for m when k = 4 and $n_1 = n_2 = n_3 = n_4 = n$.

$$\phi(m) = \frac{m^4}{n^4}\binom{n}{\frac{n+m}{2}}^4 \frac{1}{8^n} \quad \text{; So that}$$

$$\frac{\phi(m+2)}{\phi(m)} = \frac{\frac{(m+2)^4}{n^4}\binom{n}{\frac{n+m+2}{2}}^4 \frac{1}{8^n}}{\frac{m^4}{n^4}\binom{n}{\frac{n+m}{2}}^4 \frac{1}{8^n}}$$

$$= \frac{(m+2)^4}{m^4}\left(\frac{(\frac{n+m}{2})!(\frac{n-m}{2})!}{(\frac{n+m+2}{2})!(\frac{n-m-2}{2})!}\right)^4$$

Table 3: MLE for m when $k = 3$ assuming $n_1 = n_2 = n_3 = n$

n	**Sqrt(n)**	**m**
10	3.16	2
20	4.47	3
100	10	9
250	15.81	14
500	22.36	21
1000	31.63	30
10000	100	99
340	18.44	17
269	16.40	15
1056	32.50	31
1200	34.64	33
10000000	3162.28	3161

Thus, $\frac{\phi(m+2)}{\phi(m)} \geq 1$ if and only if

$$\frac{(m+2)^4}{m^4}\left(\frac{(\frac{n+m}{2})!(\frac{n-m}{2})!}{(\frac{n+m+2}{2})!(\frac{n-m-2}{2})!}\right)^4 \geq 1$$

$$\text{or} \quad \frac{(m+2)^4}{m^4} \geq \left(\frac{(\frac{n+m+2}{2})!(\frac{n-m-2}{2})!}{(\frac{n+m}{2})!(\frac{n-m}{2})!}\right)^4$$

Solving the above condition, we again see that the maximum likelihood estimate for m is about square root of n.

In Table 4,we estimate the MLE for m for some values of $n_1 = n_2 = n_3 = n_4 = n$.

For future work, we recommend to analytically prove a theorem that supports this conclusion for all values of k.

We now calculate the maximum likelihood value for m using Bayesian analysis.

4.1 Bayesian Analysis for m

Next, we move on to a case when Peter (who plays a fair coin toss game) has some money, m dollar, in his wallet. He does not remember how much money was there

Table 4: MLE for m when $k = 4$ assuming $n_1 = n_2 = n_3 = n_4 = n$

n	**Sqrt(n)**	**m**
10	3.16	2
20	4.47	3
100	10	9
250	15.81	14
500	22.36	21
1000	31.63	30
10000	100	99
340	18.44	17
269	16.40	15
1056	32.50	31
1200	34.64	33
10000000	3162.28	3161

is the wallet before he started playing the game. However, all he remembered was that the amount was less than $\$M$. Keeping this in mind, we look at the Bayesian analysis for computing m.

Let us assume that m follows a uniform distribution, that is the prior distribution for m, $f(m) = \frac{1}{M}$ where $1 \leq m \leq M$. In this case,

$$f(n|m) = \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n} \text{ ; and}$$

$$\begin{aligned} f(n,m) &= f(m) \times f(n|m) \\ &= \frac{m}{n}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n} \times \frac{1}{M} \\ &= \frac{m}{Mn}\binom{n}{\frac{n+m}{2}}\frac{1}{2^n}. \end{aligned}$$

Case 1: Let us assume $M = 20$ to show how the analysis proceeds. The highest m can be is 20. Also, if n turns out to be even, m can only take even values less than n.

$$\text{Thus, } f(n) = \frac{1}{10n2^n} \sum_{m \leq min[n,20] \ \& \ m \ is \ even} m\binom{n}{\frac{n+m}{2}}.$$

So,

$$f(m=2|n=10) = \frac{f(m=2,n=10)}{f(n=10)}$$

$$\text{So that, } f(m=2|n=10) = \frac{2\binom{10}{6}}{2\binom{10}{6}+4\binom{10}{7}+6\binom{10}{8}+8\binom{10}{9}+10\binom{10}{10}}$$

$$= \frac{420}{420+480+270+80+10}$$

$$= \frac{420}{1260},$$

$$f(m=4|n=10) = \frac{f(m=4,n=10)}{f(n=10)}$$

$$= \frac{4\binom{10}{7}}{2\binom{10}{6}+4\binom{10}{7}+6\binom{10}{8}+8\binom{10}{9}+10\binom{10}{10}}$$

$$= \frac{480}{420+480+270+80+10}$$

$$= \frac{480}{1260},$$

$$f(m=6|n=10) = \frac{f(m=6,n=10)}{f(n=10)}$$

$$= \frac{6\binom{10}{8}}{2\binom{10}{6}+4\binom{10}{7}+6\binom{10}{8}+8\binom{10}{9}+10\binom{10}{10}}$$

$$= \frac{270}{420+480+270+80+10}$$

$$= \frac{270}{1260},$$

$$f(m=8|n=10) = \frac{f(m=8,n=10)}{f(n=10)}$$

$$= \frac{8\binom{10}{9}}{2\binom{10}{6}+4\binom{10}{7}+6\binom{10}{8}+8\binom{10}{9}+10\binom{10}{10}}$$

$$= \frac{80}{420+480+270+80+10}$$

$$= \frac{80}{1260}, \text{ and}$$

$$
\begin{aligned}
f(m=10|n=10) &= \frac{f(m=10,n=10)}{f(n=10)} \\
&= \frac{10\binom{10}{10}}{2\binom{10}{6}+4\binom{10}{7}+6\binom{10}{8}+8\binom{10}{9}+10\binom{10}{10}} \\
&= \frac{10}{420+480+270+80+10} \\
&= \frac{10}{1260}.
\end{aligned}
$$

As we would expect, $\sum_{m\leq min[n,M]} f(m|n) = f(m=2|n=10) + f(m=4|n=10) + f(m=6|n=10) + f(m=6|n=10) + f(m=8|n=10) + f(m=10|n=10) = \frac{420}{1260} + \frac{480}{1260} + \frac{270}{1260} + \frac{80}{1260} + \frac{10}{1260} = 1$.

A similar analysis can be considered for other priors on m and for other values of M. If n turns out to be odd, the posterior distribution of m will take odd values no larger than $min[n, M]$.

5 Conclusions and Future Work

This thesis is an extension of the Arcsine distribution with an assumption of a lead of a strictly positive quantity $\$m$ while playing a fair coin toss game. It focuses on the distribution of the number of steps required to go bankrupt when the player started with some lead. For the first part, it was assumed that the initial lead m was known. For the later part of the thesis, m was assumed to be unknown and maximum likelihood value for m was computed. Further it was assumed that m follows a known prior distribution and posterior distribution for m was calculated.

Some of the future work may include getting better estimates of the quantiles of the FPO distribution. Secondly, this thesis looked at uniform distribution as the known prior distribution for m. As a part of the further work, Bayesian analysis can be performed with other priors too. Finally, the probability mass function for the FPO distribution may be simplified further.

BIBLIOGRAPHY

[1] C. Banderier and P. Flajolet, "Basic analytic combinatorics of directed lattice paths," *Theoretical Computer Science*, vol. 281, no. 1-2, pp. 37–80, 2002.

[2] G. Chartrand, L. Lesniak, and P. Zhang, *Graphs & digraphs*. CRC press, 2010, vol. 39.

[3] S. Connolly, Z. Gabor, and A. Godbole, "The location of the first ascent in a 123-avoiding permutation," *arXiv preprint arXiv:1401.2691*, 2014.

[4] R. L. Faber, "Differential geometry and relativity theory : an introduction," New York, 1983.

[5] W. Feller, "An introduction to probability theory and its applications," 1957.

[6] R. Martin, "The st. petersburg paradox," *Stanford Encyclopedia of Philosophy*, 2011.

[7] C. N. Moore, "Random walks," *Ramanujan Mathematical Society Mathematics Newsletter*, vol. 17, no. 3, pp. 78–84, 2007.

[8] R. Rasnick, "Generalizations of the arcsine distribution," Master's thesis, East Tenessee State University, 2019.

[9] A. Regev, "A proof of catalan's convolution formula," *Integers*, 2012.

[10] H. Robbins, "A remark on stirling's formula," *The American mathematical monthly*, vol. 62, no. 1, pp. 26–29, 1955.

[11] R. P. Stanley, *Catalan numbers*. Cambridge University Press, 2015.

VITA

ARADHANA SONI

Education:	M.S. Mathematical Sciences, East Tennessee State University, Johnson City, Tennessee 2020 B.A. (Hons.) Economics, University of Delhi, Delhi, India 2009
Professional Experience:	Teaching Assistant, East Tennessee State University, Johnson City, Tennessee, 2019–2020 Math Tutor, East Tennessee State University, Johnson City, Tennessee, 2018–2019 Assistant Actuarial Manager, Mercer India Pvt. Ltd., Gurgaon, India, 2009–2017