# The Elliptical Case of an Odds Inversion Problem

Hilmer, Kieran Jin, Angela Lycan, Ron
Ponomarenko, Vadim

May 19, 2022

**Abstract**

A recent paper by R. Moniot investigates the problem of, given a probability $\frac{p}{q}$, finding a number of red and blue balls such that, when drawing two balls without replacement, the probability of drawing different colored balls is $\frac{p}{q}$. In this paper we deepen our understanding of the case where $\frac{p}{q} > \frac{1}{2}$ by finding bounds of the number of solutions for a given probability $\frac{m}{2m-1}$ with $m \in \mathbb{N}$ and characterize "families" of probabilities that are guaranteed to have more than two solutions. We also estimate the number of achievable probabilities in the ranges $\left[\frac{m}{2m-1}, 1\right]$ and $\left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right)$. Finally, we show that the "recycling recurrence" only exists for $x_1 = n^2 - n$, $y_1 = n^2$, and $y_2 = n^2 + n$ for $n \in \mathbb{N}$.

## 1 Introduction

Drawing balls from urns is of great interest (see [3], [6], [7], [8]). A Varsity Math problem posed the question of finding a number of red and blue balls such that, when drawing two balls without replacement, the probability of drawing different colored balls is $\frac{1}{2}$ ([5]). A recent paper by R. Moniot ([4]) investigates generalizing that problem to an arbitrary probability. More specifically, it asks, given a probability $\frac{p}{q}$, is there a combination of $x$ red balls and $y$ blue balls such that the probability of drawing, without replacement, two different colored balls is $\frac{p}{q}$? The with replacement case was solved in [2]. [4] separated the probabilities into two main cases: the "elliptical case", for probabilities greater than $\frac{1}{2}$ (as solutions lie on ellipses) and the "hyperbolic case", for probabilities less than $\frac{1}{2}$ (as solutions lie on hyperbolas). In this paper, we work towards a deeper understanding of the elliptical case of the odds inversion problem.

Given $x$ red balls and $y$ blue balls, we denote the probability of drawing two different colored balls with $P(x, y)$. We will make use of much of the same background as in [4]. We must have $x, y \in \mathbb{N}$. We have

$$P(x, y) = \frac{2xy}{(x+y)(x+y-1)}. \tag{1}$$

We call $(x, y)$ a solution of a probability $\frac{p}{q}$ iff $P(x, y) = \frac{p}{q}$. A probability $\frac{p}{q}$ is called achievable iff there exist $x, y \in \mathbb{N}$ with $P(x, y) = \frac{p}{q}$.

We will assume that $0 \leq x$, $0 \leq y$, and $x + y \geq 2$. We will similarly assume that $x \leq y$, as (1) is symmetric in $x$ and $y$. Because we are only dealing with probabilities greater than $\frac{1}{2}$, we may safely assume that $x \neq 0$. We denote $t = x + y$ and $v = y - x$.

We will call a pair of solutions $(x_1, y_1), (x_2, y_2)$ a recycling recurrence iff

$$x_2 = y_1, \quad y_2 = \frac{y_1(y_1 - 1)}{x_1}. \tag{2}$$

We will usually write such a pair as $(x_1, y_1), (y_1, y_2)$. These pairs are interesting because the term $y_1$ is "recycled" and $P(x_1, y_1) = P(y_1, y_2)$.

We use the following property to motivate our next definition:

**Lemma 1.1** ([4])**.** *For $m \in \mathbb{Z}$ with $m \geq 2$, $P(m-1, m) = P(m, m) = \frac{m}{2m-1}$.*

*Proof.* Evaluating (1) with $(m-1, m)$, we have

$$P(m-1, m) = \frac{2(m-1)m}{(2m-1)(2m-2)} = \frac{m}{2m-1}.$$

Now evaluating (1) with $(m, m)$, we have

$$P(m, m) = \frac{2m^2}{(2m)(2m-1)} = \frac{m}{2m-1}.$$

$\square$

Because we are guaranteed probabilities of the form $\frac{m}{2m-1}$ for some $m \in \mathbb{N}$, investigating properties of these solutions is a natural way to study the elliptical case. We will call probabilities of the form $\frac{m}{2m-1}$ for $m \in \mathbb{N}$ "regular" and other probabilities "irregular". Note that we can write irregular probabilities as $\frac{m}{2m-1}$ for $m \in \mathbb{Q}$ with $m > 1$. Also note that for $m_1 < m_2$, $\frac{m_1}{2m_1-1} > \frac{m_2}{2m_2-1}$. We will use "balanced" to refer to solutions of the form $(m-1, m)$ and $(m, m)$ and "imbalanced" to refer to other solutions. Because we are guaranteed balanced solutions, we are primarily interested in imbalanced solutions, for both regular and irregular probabilities.

There is a recycling recurrence of the form $(m-1, m), (m, m)$ for probability $\frac{m}{2m-1}$; as we already completely understand these, when dealing with recycling recurrences we will assume they are between imbalanced solutions. We will find in section 2 that recycling recurrences between imbalanced solutions only exist for probabilities of the form $\frac{n^2}{2n^2-1}$ for some $n \in \mathbb{N}$ with $n > 1$.

We now show an important theorem that we will reference several times:

**Theorem 1.2.** *For $0 < x \leq y - 1$, $\frac{\partial P(x,y)}{\partial x}$ is strictly positive and $\frac{\partial P(x,y)}{\partial y}$ is strictly negative.*

*Proof.* Examine the partial derivatives of $P(x, y)$. We have

$$\frac{\partial P(x,y)}{\partial x} = \frac{2y(y^2 - y - x^2)}{(x+y-1)^2(x+y)^2}.$$

Because $x \leq y-1$, $x^2 < y(y-1) = y^2 - y$, and so $y^2 - y - x^2 > 0$. All other terms in the partial derivative are positive, so the partial derivative with respect to $x$ is strictly positive.

Similarly, we have

$$\frac{\partial P(x,y)}{\partial y} = \frac{2x(x^2 - x - y^2)}{(x+y-1)^2(x+y)^2}.$$

We know that $x^2 < y^2$, so $x^2 - x - y^2 < 0$, meaning the partial derivative with respect to $y$ is strictly negative. $\square$

Finally, we will show an important fact for probabilities greater than $\frac{1}{2}$.

**Theorem 1.3** ([4])**.** *Let $x, y \in \mathbb{N}$. Suppose that $P(x, y) \geq \frac{m}{2m-1}$ for some $m \in \mathbb{Z}$. Then $y \leq m$.*

*Proof.* Suppose that $P(x, y) \geq \frac{m}{2m-1}$. By Theorem 1.2, $P(x, y) \leq P(y-1, y) = \frac{y}{2y-1}$. Because $f(m) = \frac{m}{2m-1}$ is a decreasing function, we must have $y \leq m$. $\square$

**Corollary 1.3.1.** *Let $x, y \in \mathbb{N}$. Suppose that $P(x, y) \geq \frac{m}{2m-1}$ for some $m \in \mathbb{Q}$. Then $y \leq \lceil m \rceil$.*

*Proof.* Again, note that the function $f(m) = \frac{m}{2m-1}$ is strictly decreasing. Because $\lceil m \rceil \geq m$, we have $\frac{\lceil m \rceil}{2\lceil m \rceil - 1} < \frac{m}{2m-1}$. By 1.3, if $P(x, y) > \frac{\lceil m \rceil}{2\lceil m \rceil - 1}$, then $y \leq \lceil m \rceil$. Combining these facts, if $P(x, y) > \frac{m}{2m-1}$, then $y \leq \lceil m \rceil$. $\square$

This theorem allows us to enumerate all solutions for probabilities greater than $\frac{1}{2}$. For example, probability $\frac{2}{3} = \frac{2}{2(2)-1}$ has two solutions, $(1, 2)$ and $(2, 2)$. Probability $1 = \frac{1}{2(1)-1}$ has only one solution, $(1, 1)$, making $m = 1$ the sole integer $m$ value with only one solution (as $(m-1, m) = (0, 1)$ breaks our assumption that $x \neq 0$). We will revisit this property in greater detail in Section 4.

We will show that the only recycling recurrences not of the form $(m - 1, m), (m, m)$ are of the form $(n^2 - n, n^2), (n^2, n^2 + n)$ for some $n \in \mathbb{N}$. We will then show upper and lower bounds for the number of solutions for regular probabilities as well as characterizing several "families" of regular probabilities in which we are guaranteed imbalanced solutions. Finally, we will investigate the distribution of achievable probabilities in the elliptical case by finding estimates for the number of probabilities in $\left[\frac{m}{2m-1}, 1\right]$ and $\left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right)$. For example, we will show that there are $\mathcal{O}(m^{\frac{3}{2}})$ probabilities greater than or equal to a regular probability $\frac{m}{2m-1}$.

## 2 Recycling Recurrence

The recycling recurrence was of importance in the study of the odds inversion problem. We will now work towards a complete characterization of the recycling recurrence for probabilities greater than $\frac{1}{2}$.

**Theorem 2.1** ([4])**.** *There is exactly one recycling recurrence between imbalanced solutions when $m = 2n^2$ with $n \in \mathbb{N}$ and $n \geq 2$ and none otherwise.*

In order to prove this theorem, we need the following results:

**Lemma 2.2.** *Suppose that two imbalanced solutions $(x_1, y_1), (y_1, y_2)$ have probabilities greater than $\frac{1}{2}$. These solutions form a recycling recurrence iff $x_1 = n^2 - n$, $y_1 = n^2$, and $y_2 = n^2 + n$ for some $n \in \mathbb{N}$ with $n \geq 2$.*

*Proof.* ($\implies$) Suppose that $(x_1, y_1)$ and $(x_2, y_2)$ satisfy the recycling recurrence. We then know that

$$\begin{aligned} y_2 = \frac{y_1^2 - y_1}{x_1} &\implies \\ x_1 y_2 = y_1^2 - y_1. \end{aligned}$$

Hence $x_1 y_2 = y_1(y_1 - 1)$. Because we are assuming $(x_1, y_1)$ and $(x_2, y_2)$ are imbalanced, we know that $x_1 \neq y_1 - 1$ and $x_1 \neq y_1$. Thus, writing $y_1 = a_1 a_2$ and $y_1 - 1 = b_1 b_2$ for $a_1, b_1, a_2, b_2 \in \mathbb{N}$, we have $x_1 = a_1 b_1$ and $y_2 = a_2 b_2$ or vice versa. We aim to show that $a_1 = a_2$. Essentially, we will bound $|a_2 - b_1|$ and $|b_2 - a_1|$, thereby reducing the problem to a finite number of cases. We can then check those cases to see that $a_1 = a_2$. WLOG, we may assume that $a_1 \leq a_2$. Then let $x_1$ equal the smaller of $a_1 b_1$ and $a_2 b_2$, and $y_2$ the larger. We immediately see that $a_1 \neq b_2$ and $a_2 \neq b_1$, or else we would violate our assumption that $(x_1, y_1)$ and $(x_2, y_2)$ are imbalanced.

We will first show that $|b_2 - a_1| = 1$. For any $(x, y)$ with probability greater than $\frac{1}{2}$, we have

$$\begin{aligned} \frac{t^2 - v^2}{2t(t-1)} > \frac{1}{2} &\implies \\ t^2 - v^2 > t^2 - t &\implies \\ v^2 < t. \end{aligned} \tag{3}$$

If $a_2 b_2 = x_1$, then the first solution is $(a_2 b_2, a_1 a_2)$. If, on the other hand, $a_2 b_2 = y_2$, then the second solution is $(a_1 a_2, a_2 b_2)$. The $t$ values for these two solutions are the same, and the $v$ values are additive inverses. Because inequality (3) involves $v^2$, the inequality is the same in either case. Substituting the $t$ and $v$ values into (3) gives

$$\begin{aligned} (a_2 b_2 - a_1 a_2)^2 < a_2 b_2 + a_1 a_2 &\implies \\ a_2 (b_2 - a_1)^2 < b_2 + a_1 &\implies \\ a_2 < \frac{b_2 + a_1}{(b_2 - a_1)^2}. \end{aligned} \tag{4}$$

We also have $a_1 \le a_2$, so

$$a_1 < \frac{b_2 + a_1}{(b_2 - a_1)^2} \implies$$
$$a_1((b_2 - a_1)^2 - 1) < b_2. \tag{5}$$

Let $k = b_2 - a_1$. Then

$$a_1(k^2 - 1) < b_2 = a_1 + k \implies$$
$$a_1(k^2 - 2) < k.$$

Suppose by way of contradiction that $|k| > 2$. Then $k^2 > 4$, so $k^2 - 2 > 2 > 0$, and we have $a_1 < \frac{k}{k^2-2}$. We can see that $k > 0$, or else $a_1 < 0$. We now see that $k^2 > 2k > k + 2$, so $k^2 - 2 > k$. Hence $\frac{k}{k^2-2}$ is strictly less than 1. But this is impossible, since $a_1$ is a positive integer. As such, $-2 \le b_2 - a_1 \le 2$.

Now suppose $k = 2$. Then $3a_1 = a_1(k^2 - 1) < b_2 = 2 + a_1$, meaning $a_1 < 1$, which is impossible. Similarly, if $b_2 - a_1 = -2$, then $a_1 < -1$, which is also impossible. Hence $-2 < b_2 - a_1 < 2$. But $b_2 - a_1$ is a nonzero integer, so it must equal either 1 or -1.

We will now show that $|a_2 - b_1| = 1$. By the construction of $a_1, a_2, b_1, b_2$, either $(x_1, y_1) = (a_2b_2, a_1a_2)$ or $(y_1, y_2) = (a_1a_2, a_2b_2)$. Since $a_1a_2 = y_1$ and $b_1b_2 = y_1 - 1$, $a_1a_2 = b_1b_2 + 1$. Applying a similar argument as to how we last calculated $v$ and $t$, we can substitute to see that $v^2 = (a_2b_2 - b_1b_2 - 1)^2$ and $t = a_2b_2 + b_1b_2 + 1$. Applying (3) to these new values gives

$$(a_2b_2 - b_1b_2 - 1)^2 < a_2b_2 + b_1b_2 + 1 \implies$$
$$(a_2b_2 - b_1b_2)^2 - 2a_2b_2 - 2b_1b_2 < a_2b_2 + b_1b_2 \implies$$
$$b_2^2(a_2 - b_1)^2 < 3a_2b_2 - b_1b_2 \implies$$
$$b_2(a_2 - b_1)^2 < 3a_2 - b_1 \implies$$
$$b_2 < \frac{3a_2 - b_1}{(a_2 - b_1)^2}. \tag{6}$$

We have shown that $(b_2 - a_1)^2 = 1$. Substituting that into (4) gives

$$a_2 < b_2 + a_1 \implies$$
$$a_2 - b_2 < a_1 \implies$$
$$a_1a_2 > a_2^2 - a_2b_2$$

Recall that $a_1a_2 = b_1b_2 + 1$:

$$b_1b_2 + 1 > a_2^2 - a_2b_2 \implies$$
$$b_1b_2 + a_2b_2 > a_2^2 - 1 \implies$$
$$(b_1 + a_2)b_2 > a_2^2 - 1 \implies$$
$$b_2 > \frac{a_2^2 - 1}{b_1 + a_2}. \tag{7}$$

Combining (7) and (6) gives

$$\begin{aligned}
\frac{a_2^2-1}{b_1+a_2} < \frac{3a_2-b_1}{(a_2-b_1)^2} &\implies \\
(a_2^2-1)(a_2-b_1)^2 < (3a_2-b_1)(a_2+b_1) &\implies \\
a_2^2-4-2a_2b_1+b_1^2 < 0 &\implies \\
(b_1-a_2)^2 < 4 &\implies \\
|b_1-a_2| < 2.&
\end{aligned}$$

Since $a_2 \neq b_1$, $b_1 - a_2$ must also be either 1 or -1.

We now know that $|b_1 - a_2| = 1 = |b_2 - a_1|$. If $b_1 - a_2 = 1 = b_2 - a_1$, then we have $b_1 = a_2 + 1$ and $b_2 = a_1 + 1$. Then $y_1 - 1 = b_1 b_2 = (a_1+1)(a_2+1) = y_1 + a_1 + a_2 + 1 > y_1$, which is a contradiction. If $b_1 - a_2 = -1 = b_2 - a_1$, then $b_1 = a_2 - 1$ and $b_2 = a_1 - 1$. Then $y_1 - 1 = (a_1-1)(a_2-1) = y_1 - a_1 - a_2 + 1$. We thus have $2 = a_1 + a_2$, so $a_1 = a_2 = 1$. This implies that $y_1 = 1$, but there are no recycling recurrences with $y_1 = 1$, so we have a contradiction. We now know that one of $b_1 - a_2$ and $b_2 - a_1$ equals 1 and the other equals $-1$. Assume that $b_2 = a_1 + 1$ and $b_1 = a_2 - 1$. Then

$$a_1a_2 - 1 = y_1 - 1 = (a_1+1)(a_2-1) = a_1a_2 + a_2 - a_1 - 1,$$

so $0 = a_2 - a_1$, or $a_1 = a_2$. If $b_2 = a_1 - 1$ and $b_1 = a_2 + 1$, then we get $y_1 - 1 = y_1 - a_2 + a_1 - 1$, so still $a_1 = a_2$. Now that $x_1 = b_1 a_1$ and $y_2 = b_2 a_1$, we can see that $b_1 < a_1 < b_2$. Thus, $b_1 + 1 = a_1 = a_2 = b_2 - 1$. Some simple algebra will now show that $x_1 = n^2 - n$, $y_1 = n^2$, and $y_2 = n^2 + n$ for some $n \in \mathbb{N}$ (particularly $n = a_1$). Note, however, that $n \neq 1$ or else $x_1 = 0$, which contradicts our assumption that $x \neq 0$.

( $\Longleftarrow$ ) Suppose that $x_1 = n^2 - n$, $y_1 = n^2$, and $y_2 = n^2 + n$. Simple algebra will verify that $(x_1, y_1), (y_1, y_2)$ satisfies the recycling recurrence. $\square$

There are many probabilities in the hyperbolic case, $P(x,y) < \frac{1}{2}$, which have three solutions of the form $(x_1,y_1),(y_1,y_2),(y_2,y_3)$, i.e. where the second term of one recycling recurrence is also the first term of a different recycling recurrence. We will now show that these "recycling triples" are entirely absent in the elliptical case.

**Corollary 2.2.1.** *There are no recycling triples.*

*Proof.* Applying Lemma 2.2, we see that if $(x_1, y_1)$, $(x_2, y_2)$ is a recycling recurrence, we must have $x_1 = n^2 - n$, $y_1 = n^2$, and $y_2 = n^2 + n = n(n+1)$. We can see that $\nexists z \in \mathbb{N} \mid y_2 = z^2$. Applying Lemma 2.2 again, we cannot have a recycling pair $(x_2, y_2), (x_3, y_3)$. Thus, there are no recycling triples. $\square$

**Lemma 2.3.** *$(x_1, y_1)$ and $(x_2, y_2)$ are imbalanced solutions that satisfy the recycling recurrence giving probability $\frac{m}{2m-1}$ iff $m = 2n^2$, $y_1 = n^2$, and $x_1 = n^2 - n$ for some $n \in \mathbb{N}$ with $n \geq 2$.*

*Proof.* ( $\Longrightarrow$ ) Suppose that $(x_1, y_1)$ and $(x_2, y_2)$ satisfy the recycling recurrence with probability $\frac{m}{2m-1}$. Because $(x_1, y_1)$ and $(x_2, y_2)$ satisfy the recycling recurrence, we can apply Lemma 2.2 to see that $x_1 = n^2 - n$ and $y_1 = n^2$ for some $n \in \mathbb{N}$ with $n \geq 2$. We now substitute:

$$\begin{aligned}
\frac{2xy}{(x+y)(x+y-1)} &= \\
\frac{2(n^2-n)(n^2)}{(n^2-n+n^2-1)(n^2-n+n^2)} &= \\
\frac{2(n^4-n^3)}{(2n^2-n)(2n^2-n-1)} &= \\
\frac{2(n^4-n^3)}{4n^4-4n^3-n^2+n} &= \\
\frac{2n^2(n-1)}{(n-1)(2n-1)(2n+1)} &= \\
\frac{2n^2}{(2n-1)(2n+1)} &= \\
\frac{2n^2}{4n^2-1} &= \\
\frac{m}{2m-1}&.
\end{aligned}$$

We can see that $m = 2n^2$ with $n \geq 2$, as desired.

( $\Longleftarrow$ ) Suppose that $m = 2n^2$, $y_1 = n^2$, and $x_1 = n^2 - n$ for some $n \in \mathbb{N}$ and $n \geq 2$. Then $(n^2, n^2 + n)$ satisfies the recycling recurrence. $\square$

Note that these solutions occur at $y_1 = \frac{m}{2}$—from a geometric perspective, they occur near the center of the ellipse for probability $\frac{m}{2m-1}$. We are now ready to prove Theorem 2.1.

*Proof.* (there is a RR when $m = 2n^2$): Apply Lemma 2.3.

(there is exactly one RR when $m = 2n^2$): Apply Corollary 2.2.1 to see that there are no recycling triples. Applying Lemma 2.3, we see that there is a recycling recurrence iff $y = \frac{m}{2}$, implying that there aren't distinct pairs of values satisfying the RR.

(the RR doesn't happen otherwise): Apply Lemma 2.3. $\square$

Combined with our knowledge of recycling recurrences of the form $(m-1, m), (m, m)$, Theorem 2.1 provides a complete characterization of the recycling recurrence for probabilities greater than $\frac{1}{2}$.

# 3 Solutions for Regular Probabilities

We will now explore solutions for regular probabilities; in particular, we will determine how many solutions exist for these probabilities and characterize the imbalanced solutions for these probabilities.

## 3.1 Upper Bound on Number of Solutions

Let $n \in \mathbb{N}$ with prime factorization $n = p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m}$. Recall that the prime omega function is defined $\omega(n) = m$. We have the following conjecture for an upper bound on the number of solutions for $P(x, y) = \frac{m}{2m-1}$.

**Conjecture 1.** *There are at most $2^k$ solutions giving probability $\frac{m}{2m-1}$, where $k = \omega(2m-1)$.*

This conjecture is true for $m \leq 3000$. Additionally, we have a partial proof of the conjecture. First, we need two results, which we will also use in Section 4. We will first introduce the following lemma:

**Lemma 3.1.** *For $m > y > 1$, $P(x, y) = \frac{m}{2m-1}$ iff*

$$x = \frac{m - 2y + 2my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}.$$

*Proof.* ($\Longleftarrow$) Suppose that $(x, y)$ is a solution to $\frac{m}{2m-1}$ i.e.

$$\begin{aligned}
\frac{2xy}{(x+y)(x+y-1)} &= \frac{m}{2m-1} \implies \\
(2xy)(2m-1) &= m((x+y)(x+y-1)) \implies \\
-mx^2 + 2mxy + mx - 2xy - my^2 + my &= 0 \implies \\
-mx^2 + (2my + m - 2y)x - (y-1)my &= 0 \implies \\
x = \frac{(2my + m - 2y) \pm \sqrt{(2my + m - 2y)^2 - 4m^2 y(y-1)}}{2m} & \implies \\
x = \frac{m - 2y + 2my \pm m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}&.
\end{aligned}$$

One can verify that, for $y < m$,

$$y < \frac{m - 2y + 2my + m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}.$$

Because we are assuming $x \leq y$, we can replace the $\pm$ with $-$. Finally,

$$x = \frac{m - 2y + 2my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}.$$

($\Longrightarrow$) Let

$$x = \frac{m - 2y + 2my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}.$$

We will now show that $P(x,y) = \frac{m}{2m-1}$. We have

$$\begin{aligned}
2xy &= \frac{y(m - 2y + 2my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}})}{m} \\
x + y &= \frac{m - 2y + 4my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m} \\
x + y - 1 &= \frac{-m - 2y + 4my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}.
\end{aligned}$$

Thus,

$$(x+y)(x+y-1) = \frac{y(2m-1)(2my + m - 2y - m)\sqrt{1 + \frac{4(2m-1)y(m-y)}{m^2}}}{m^2}.$$

And finally,

$$\frac{2xy}{(x+y)(x+y-1)} = \frac{m}{2m-1}.$$

□

Let

$$\tilde{x}(m,y) = \frac{m - 2y + 2my - m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m}. \tag{8}$$

We have the following corollary:

**Corollary 3.1.1.** *For $y > 1$, $P(x,y) > \frac{m}{2m-1}$ iff*

$$x \geq \tilde{x}(m,y).$$

*Proof.* By Theorem 1.2, $\frac{\partial P(x,y)}{\partial x}$ is strictly positive. We can combine this fact with Lemma 3.1 to see that Corollary 3.1.1 is true for $y \neq m$. When $y = m$, one can verify that $\tilde{x}(m,y) = m - 1$ and

$$\frac{m - 2y + 2my + m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m} = m.$$

By Lemma 1.1, both $(m-1, m)$ and $(m,m)$ are solutions of $\frac{m}{2m-1}$ for $m \neq 1$. Because $m > m - 1$, we must still have $x \geq m - 1$. Thus, the theorem is also true for $y = m$. □

Now, let

$$z(y,m) = \sqrt{m^2 + 8m^2y + 4y^2 - 4ym(1+2y)}. \tag{9}$$

In order to move towards an upper bound on the number of solutions, we will establish the following condition.

**Lemma 3.2.** *For a given $y, m \in \mathbb{N}$, if there exists an $x \in \mathbb{N}$ such that $(x, y)$ is a valid pair of balls with $P(x, y) = \frac{m}{2m-1}$ then $z(y, m) \in \mathbb{Z}$.*

*Proof.* Suppose that there exists an $x \in \mathbb{N}$ such $P(x, y) = \frac{m}{2m-1}$. Then we know that

$$\begin{aligned} x &= \frac{m - 2y + 2my - z(y,m)}{2m} \implies \\ 2mx &= m - 2y + 2my - z(y,m). \end{aligned}$$

Because $2mx \in \mathbb{N}$, the RHS must be an integer as well. Because $m, 2y, 2my \in \mathbb{Z}$, we must have $z(y, m) \in \mathbb{Z}$ as well. $\square$

We can now provide a partial proof of Conjecture 1.

*Partial Proof.* Consider the necessary condition $z(y, m) \in \mathbb{N}$.

$$\begin{aligned} \sqrt{m^2 + 8m^2y + 4y^2 - 4ym(1 + 2y)} &= z(y,m) \implies \\ m^2 - 4ym + 8m^2y + 4y^2 - 8my^2 &= z(y,m)^2 \implies \\ 2m^3 - 8m^2y - m^2 + 8my^2 + 4my - 4y^2 &= 2m^3 - z(y,m)^2 \implies \\ y^2 - my + \frac{m^2}{4} &= \frac{2m^3 - z(y,m)^2}{4(2m-1)} \implies \\ (y - \frac{m}{2}) &= \sqrt{\frac{2m^3 - z(y,m)^2}{4(2m-1)}} \implies \\ 2y &= \sqrt{\frac{2m^3 - z(y,m)^2}{2m-1}} + m. \end{aligned}$$

Because $2y \in \mathbb{Z}$, we must have $\sqrt{\frac{2m^3 - z(y,m)^2}{2m-1}} + m \in \mathbb{Z}$. Thus, we must have $\sqrt{\frac{2m^3 - z(y,m)^2}{2m-1}} \in \mathbb{Z}$, and so necessarily $\frac{2m^3 - z(y,m)^2}{2m-1} \in \mathbb{Z}$. As such, we must have $z(y, m)^2 \equiv_{2m-1} 2m^3$, which means that

$$z(y,m)^2 \equiv_{2m-1} 2m^3 - m^2(2m-1) \equiv_{2m-1} m^2.$$

To find values of $z(y, m)$, we would apply the Chinese Remainder Theorem. For more information on this process, see [1]. There are at most $2^k$ ways to do this, where $k = \omega(2m - 1)$. Now suppose that each value of $z(y, m)$ corresponds to at most two solutions $(x, y)$. For each of $y = 0$ and $y = m$, there are two solutions giving $\frac{m}{2m-1}$. Thus, the number of solutions found this way is $2(2^k) + 2 = 2^{k+1} + 2$. Remove the "trivial solutions" $(0, 0), (0, 1)$, and $(1, 0)$ (see [4]) and the solution $(m, m)$ to get $2^{k+1} - 2$ solutions. Half of the solutions counted by this value have $x < y$ and the remaining solutions have $y < x$. Imposing the condition that $x \leq y$, we have $2^k - 1$ solutions. Add back the solution $(m, m)$ to find that there is a maximum of $2^k$ solutions.

## 3.2 Conditions for the Existence of Imbalanced Solutions

We will now show a method of verifying whether a regular probability has imbalanced solutions. In doing so, it will be useful to represent solutions using the auxiliary variables $t$ and $v$. Given a solution $(x, y)$ to a probability, we let $[t, v]$ represent the pair of integers $t$ and $v$ such that $t = x + y$ and $v = y - x$.

**Lemma 3.3.** *The regular probability $\frac{m}{2m-1}$ has an imbalanced solution $[t, v]$ iff $v^2 = \frac{t(2m-t)}{2m-1}$.*

*Proof.* Suppose that the regular probability $\frac{m}{2m-1}$ has an imbalanced solution determined by $[t, v]$. We can then substitute the expressions for $t$ and $v$ into equation (1) to get

$$\begin{aligned}
\frac{m}{2m-1} &= \frac{t^2 - v^2}{2t(t-1)} && \iff \\
2mt(t-1) &= (t^2 - v^2)(2m-1) && \iff \\
2mt^2 - 2mt &= 2mt^2 - t^2 - (2m-1)v^2 && \iff \\
(2m-1)v^2 &= 2mt - t^2 && \iff \\
v^2 &= \frac{t(2m-t)}{2m-1}. && (10)
\end{aligned}$$

Hence the solution $[t, v]$ satisfies equation (1) iff $v^2 = \frac{t(2m-t)}{2m-1}$. □

Note that, by the symmetry of the ellipse, if $(x, y)$ is a solution giving $P(x, y) = \frac{m}{2m-1}$, then $(m - y, m - x)$ is also such a solution. In terms of $t$ and $v$, the equivalent statement is that if $[t, v]$ is a solution to a given regular probability, then $[2m - t, v]$ is also such a solution, which can be readily verified using the above lemma.

**Theorem 3.4.** *If $2m - 1$ is a prime power, then the regular probability $\frac{m}{2m-1}$ has no imbalanced solutions.*

*Proof.* Suppose that $2m - 1 = p^n$ for some prime $p$ and some natural number $n$. By way of contradiction, suppose also that $\frac{m}{2m-1}$ has an imbalanced solution $[t, v]$. Note that for all imbalanced solutions, $t > 1$ and $2m - t > 1$, which in turn implies that both $t$ and $2m - t$ are both strictly less than $2m - 1 = p^n$. Now by Lemma 3.3, $\frac{t(2m-t)}{2m-1}$ is an integer. It follows that $p \mid t$ and $p \mid 2m - t$. But then $p \mid t + (2m - t) = 2m$, which contradicts $p \mid 2m - 1$. Hence $\frac{m}{2m-1}$ has no imbalanced solutions. □

As a corollary to the above theorem, we see that if $\frac{m}{2m-1}$ has imbalanced solutions, we may write $2m - 1 = ab$ where $a, b \in \mathbb{N} \setminus \{0, 1\}$ and $\gcd(a, b) = 1$.

Given a pair of natural numbers $(a, b)$, we define $a'$ and $b'$ to be the inverse of $a \pmod{b}$ and the inverse of $b \pmod{a}$ respectively. Note that in order for the modular inverses to exist, we must have $\gcd(a, b) = 1$. We will call $(a', b')$ the derived pair of $(a, b)$. We note the following property of derived pairs:

**Lemma 3.5.** *For $a, b \in \mathbb{N}$ with derived pair $(a', b')$, $aa' + bb' - 1 = ab$.*

*Proof.* By definition, $aa' \equiv_b 1$. Let $n$ be the least natural number such that $aa' = bn + 1$. Rearranging gives $-bn = -aa' + 1 \equiv_a 1$. Now $-n \equiv_a a - n$, so we also have $b(a-n) \equiv_a 1$. Since $a - n < a$, we have $b' = a - n$ or $n = a - b'$. Substituting gives $aa' = b(a - b') + 1 \Rightarrow aa' + bb' - 1 = ab$. $\square$

**Lemma 3.6.** *Let $2m-1 = ab$, and let $(a', b')$ be the derived pair of $(a, b)$. Then $\frac{t(2m-t)}{2m-1} = a'b'$ if either $t = aa'$ or $t = bb'$.*

*Proof.* We first note that $2m - 1 = ab = aa' + bb' - 1$ by Lemma 3.5. WLOG, assume that $t = aa'$. Note that $a \mid t$, and since $a' < b$, $t < 2m - 1$ as required. We now have $2m - 1 = t + bb' - 1$ or $2m - t = bb'$. Finally, we see that $\frac{t(2m-t)}{2m-1} = \frac{(aa')(bb')}{ab} = a'b'$. $\square$

We are now ready to describe some integers $m$ such that $\frac{m}{2m-1}$ has imbalanced solutions.

**Theorem 3.7.** *A regular probability $\frac{m}{2m-1}$ has at least two imbalanced solutions if there exist $a, b \in \mathbb{N} \setminus \{0, 1\}$ with derived pair $(a', b')$ such that*

- $ab = 2m - 1$
- $gcd(a, b) = 1$
- *$a'b'$ is a perfect square greater than 1.*

*Proof.* Suppose that for a regular probability $\frac{m}{2m-1}$, there exist $a$ and $b$ which satisfy the above conditions. Let $t = aa'$ or $t = bb'$, and let $v = \sqrt{a'b'}$. By Lemma 3.6, we have $\frac{t(2m-t)}{2m-1} = a'b' = v^2$. By Lemma 3.3, $[aa', a'b']$ and $[bb', a'b']$ both determine solutions for the given regular probability. However, since $x = \frac{t-v}{2}$ and $y = \frac{t+v}{2}$, we must now show that our choices of $t$ and $v$ have the same parity. As in Lemma 3.6, we may assume WLOG that $t = aa'$. We note that since $2m - 1$ is odd, both $a$ and $b$ must also be odd, and that the parity of $v$ matches the parity of $v^2 = a'b'$. If $a'$ is even, then both $t$ and $v$ are even, and we are done. If $a'$ is odd, then $t$ is also odd. Suppose by way of contradiction that $b'$ is even. Using the definition of $b'$, we now have $bb' = ak + 1$, where $k$ is an odd integer in the interval $[0, b)$. This implies

$$\begin{aligned} ak &\equiv_b -1 \implies \\ ab - ak = a(b-k) &\equiv_b 1 \implies \\ a' &= b - k, \end{aligned}$$

where in the final step we use $0 \le b - k < b$. But $a'$ is odd, while $b - k$ is even, which is a contradiction. Hence $b'$, and therefore $v$, must be odd. $\square$

**Corollary 3.7.1.** *Let $2m - 1 = ab$ as in Lemma 3.6. Then the imbalanced solutions for $\frac{m}{2m-1}$ given by $t = aa'$ and $t = bb'$ are distinct.*

*Proof.* If $t = aa'$, then the corresponding $(x, y)$ solution is $(\frac{aa'-\sqrt{a'b'}}{2}, \frac{aa'+\sqrt{a'b'}}{2})$. If instead $t = bb'$, then the $(x, y)$ solution is $(\frac{bb'-\sqrt{a'b'}}{2}, \frac{bb'+\sqrt{a'b'}}{2})$. Suppose that these two solutions are the same. Then $\frac{aa'-\sqrt{a'b'}}{2} = \frac{bb'-\sqrt{a'b'}}{2}$, and $aa' = bb'$. Reducing this mod $a$ gives $0 \equiv_a bb'$. But $bb' \equiv_a 1$. By contradiction, the two $(x, y)$ solutions given above are distinct. □

Based on Theorem 3.7, we pose the following conjecture:

**Conjecture 2.** *All solutions to a regular probability $\frac{m}{2m-1}$, including the two balanced solutions, correspond to a unique factorization $ab$ of $2m-1$, with $a, b \in \mathbb{N} \setminus \{0\}$ and $gcd(a, b) = 1$.*

Note that Conjecture 2 implies Conjecture 1. Indeed, let $p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m}$ be the prime factorization of $2m-1$. Because $\gcd(a, b) = 1$, the number of possible factorizations $ab$ of $2m-1$ corresponds to the number of ways to choose the $p_i^{\alpha_i}$ that divide $a$, which is $2^k$. Then if each solution for $\frac{m}{2m-1}$ corresponds to only one of the $2^k$ possible factorizations, the maximum number of such solutions is $2^k$.

## 3.3 Solution Families for Regular Probabilities

We will now use Theorem 3.7 to find several families of imbalanced solutions. The following result provides several conditions on $a$ and $b$ which guarantee that $a'b'$ is square.

**Theorem 3.8.** *Let $a, b \in \mathbb{N} \setminus \{0, 1\}$ have derived pair $(a', b')$. Then $a'b'$ is a perfect square greater than one if at least one of the following conditions is met:*

1. $b = a + 2$
2. $b = a(a-1) - 1$
3. $a = 6n + 1$ *and* $b = 2a + 3$ *for some* $n \in \mathbb{N}$ *with* $n \geq 1$ *or*
4. $b = (an \pm 1)^2 - an^2$ *for some* $n \in \mathbb{N}$ *with* $n \geq 1$.

*Proof.* Let $a, b \in \mathbb{N} \setminus \{0, 1\}$ with derived pair $(a', b')$.

1. Suppose that $b = a + 2$. Then $a \equiv_b a - b = -2$, and therefore for any $n \in \mathbb{N}$, $an \equiv_b -2n$. Now let $n = \frac{a+1}{2}$. Then

$$an \equiv_b -2n = -a - 1 = (b - a) - 1 - b = 1 - b \equiv_b 1.$$

So $n = a'$. We now notice that $b \equiv_a b - a = 2$, and that $bn \equiv_a 2n$ for any $n \in \mathbb{N}$. As before, we let $n = \frac{a+1}{2}$, and we have

$$bn \equiv_a 2n = a + 1 = b - 1 = (b - a) + a - 1 = 1 + a \equiv_a 1.$$

So $n = b'$, and $a'b' = n^2$ is a perfect square greater than 1.

2. Now suppose that $b = a(a-1)-1$. Then since $a(a-1) \equiv_b 1$ and $a-1 < b$, $a-1 = a'$. Now

   $$b(a-1) = ab - b = ab - a(a-1) + 1 \equiv_a 1.$$

   Clearly $a - 1 < a$, so $a - 1 = b'$, and $a'b' = (a-1)^2$ is a perfect square greater than 1.

3. Suppose now that $a = 6n+1$ and $b = 2a+3$ for some $n \in \mathbb{N}$ with $n \geq 1$. Then $b = 12n+5$. Now

   $$a(4n+1) = (6n+1)(4n+1) = 24n^2+10n+1 = 2n(12n+5)+1 = 2nb+1,$$

   which is congruent to 1 (mod $b$). Since $4n+1 < b$, $a' = 4n+1$. We also have

   $$b(4n+1) = (2a+3)(4n+1) = 8an+12n+2a+3 = a(8n+2)+2(6n+1)+1.$$

   Since $a = 6n+1$, taking this mod $a$ gives 1. Certainly $4n+1 < a = 6n+1$, so $b' = 4n+1$, and $a'b' = (4n+1)^2$ is a perfect square greater than 1.

4. Finally, suppose that $b = (an \pm 1)^2 - an^2$ for some $n \in \mathbb{N}$ with $n \geq 1$. Then

   $$b = a^2n^2 \pm 2an + 1 - an^2 = an^2(a-1) \pm 2an + 1.$$

   So $b \equiv_a 1$, and $b' = 1$. Now let $c = \big((a-1)n \pm 1\big)^2$. Note that $c < b$, since

   $$\begin{aligned} c = n^2(a-1)^2 \pm 2n(a-1) + 1 &< \\ a^2n^2 \pm 2an + 1 - an^2 &= \\ an^2(a-1) \pm 2an + 1 &= b. \end{aligned}$$

   We also have

   $$\begin{aligned} ac &= a\big((a-1)^2n^2 \pm 2(a-1)n + 1\big) = a(a-1)^2n^2 \pm 2a(a-1)n + a \\ &= (a-1)\big(a(a-1)n^2 \pm 2an + 1\big) + 1 = (a-1)b + 1 \equiv_b 1. \end{aligned}$$

   Hence $c = a'$, and $a'b' = \big((a-1)n \pm 1\big)^2$ is a perfect square greater than 1.

□

In conjunction with Theorem 3.7, the conditions in the above theorem each produce infinitely many solutions for various regular probabilities. We will refer to the sets of solutions produced by each of the conditions as "families" of solutions. We say that solutions produced by the first condition in Theorem 3.8 belong to family 1, and so on.

**Remark.** *The solutions produced by family 1 satisfy the recycling recurrence.*

*Proof.* Suppose that a regular probability $\frac{m}{2m-1}$ has a solution in family 1 given by $[t, v]$. We see that there exists an $n \in \mathbb{N} \setminus \{0, 1\}$ such that $v = n$ and $t = 2n^2 \pm n$. If $t = 2n^2 + n$, then $(x, y) = (n^2, n^2 + n)$. If $t = 2n^2 - n$, then $(x, y) = (n^2 - n, n^2)$. By Theorem 2.1, these are exactly the $(x, y)$ solutions which satisfy the recycling recurrence in the elliptical case. $\square$

We now turn our attention to the intersections between these solution families; that is, imbalanced solutions that are produced by more than one of the families. For a given regular probability, there may be multiple distinct imbalanced solutions which each belong to different families, or there may be individual imbalanced solutions which belong to multiple families simultaneously. It is simpler to analyze the latter case, because we can equate the expressions relating $a$ and $b$ for the two families we are interested in.

For families 1 and 2, we have $a + 2 = a(a - 1) - 1$. Solving for $a$ gives $a = 3$. Hence $b = 5$, and we arrive at the solutions $(2, 4), (4, 6)$ when $m = 8$. For families 1 and 3, we have $a + 2 = 2a + 3$. Solving for $a$ gives $a = -1$, which does not give a solution because $a$ must be greater than 1. For families 2 and 3, the relevant equation is $a(a - 1) - 1 = 2a + 3$. Solving gives $a = 4$, which is even and so not a possible factor of the odd number $2m - 1$.

Family 4 always gives $b' = 1$, while the other three have $a' = b'$. If a solution with $b' = 1$ were to coincide with a solution where $a' = b'$, then $a'$, and therefore $v$, would also be 1. This would produce a balanced solution, which we may disregard. So we see that an imbalanced solution for Family 4 can never belong to any of the other families.

Although $(2, 4)$ and $(4, 6)$ are the only two individual solutions that belong to more than one of the families, there are additional regular probabilities which have solutions belonging to multiple families. For example, the probability corresponding to $m = 1008$ has two solutions belonging to family 2, and two more solutions belonging to family 3. In addition, the probability corresponding to $m = 1190$ has two solutions belonging to family 4 with $n = 1$, and two more solutions also belonging to family 4, but with $n = 11$.

A computer search was conducted to find solutions for probabilities of the form $\frac{m}{2m-1}$ for all integers $m \leq 3000$. A separate search identified those solutions belonging to each of the above families. As shown in Figure 1, Family 4 covers the most solutions. There are 624 imbalanced solutions in this range which do not belong to any family, i.e. are not covered by the families.

# 4 Estimations of the Density of Achievable Probabilities

Finally, we will find estimations of the number of solutions for probabilities in the ranges $\left[\frac{m}{2m-1}, 1\right]$ and $\left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right)$ in order to better understand the density of solutions for probabilities in the range $(\frac{1}{2}, 1]$.

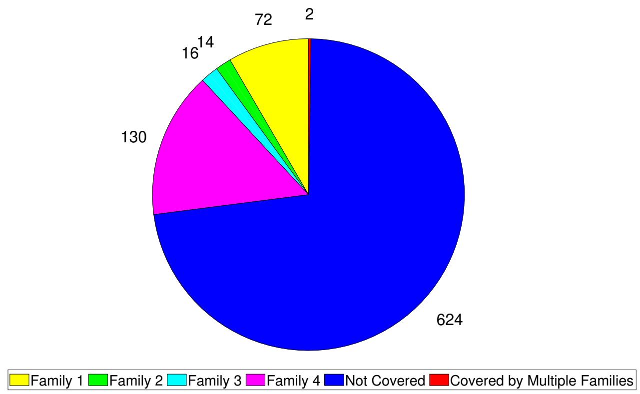


Figure 1: Number of solutions covered by each family.

## 4.1 Number of Solutions for Probabilities Greater Than or Equal to $\frac{m}{2m-1}$

We can leverage Corollary 3.1.1 to find bounds on the number of solutions for probabilities in various ranges. Let

$$S(m) = 1 + \sum_{i=2}^{m} (i + 1 - \lceil \tilde{x}(m, i) \rceil) \tag{11}$$

where $\tilde{x}$ was defined in (8).

**Lemma 4.1.** *There are $S(m)$ solutions for probabilities greater than or equal to $\frac{m}{2m-1}$.*

*Proof.* By Theorem 1.3, we must have $x \le y \le m$. We know that if $y = 1$ then there is only one $x$ value ($x = 1$) such that $(x, y)$ is a valid pair of balls. For any other fixed $y$, $\tilde{x}(m, y)$ is the value of $x$ such that $P(x, y) = \frac{m}{2m-1}$. By Corollary 3.1.1, if $x \ge \tilde{x}(m, y)$, then $P(x, y) \ge \frac{m}{2m-1}$. The least integer value in this range is $\lceil \tilde{x}(m, y) \rceil$, while the greatest is $m$. Summing over all values of $y$, we have $S(m)$ solutions for probabilities greater than or equal to $\frac{m}{2m-1}$. $\square$

We will now find an estimate of how many values are in this range. First, we need an additional lemma:

**Lemma 4.2.** *For real numbers $y, m$, we have*

$$\int_{\alpha}^{\beta} \frac{m\sqrt{1 + 8y + \frac{4y^2}{m^2} - \frac{4y(1+2y)}{m}}}{2m} dy = m\sqrt{2m-1}(\tilde{i}(\beta, m) - \tilde{i}(\alpha, m)),$$

*where*

$$\tilde{i}(c,m) = \frac{1}{8}\left((c-1)\sqrt{\frac{m}{2m-1}-(1-c)^2} - \frac{m}{2m-1}\left(\tan^{-1}\left(\frac{1-c}{\sqrt{\frac{m}{2m-1}-(1-c)^2}}\right)\right)\right)$$

*Proof.* We have

$$\int_\alpha^\beta \frac{m\sqrt{1+8y+\frac{4y^2}{m^2}-\frac{4y(1+2y)}{m}}}{2m}dy = \int_\alpha^\beta \frac{1}{2}\sqrt{1+8y+4(\frac{y}{m})^2-4\frac{y}{m}-8m(\frac{y}{m})^2}dy.$$

Let $t=\frac{y}{m}$. Then $mdt = dy$. We have

$$\int_{\frac{\alpha}{2}}^{\frac{\beta}{2}} \frac{m}{2}\sqrt{1+8mt+4t^2-4t-8mt^2}dt =$$
$$\int_{\frac{\alpha}{2}}^{\frac{\beta}{2}} m\sqrt{2m-1}\sqrt{\frac{1}{4(2m-1)}+t-t^2}dt =$$
$$m\sqrt{2m-1}\int_{\frac{\alpha}{2}}^{\frac{\beta}{2}}\sqrt{\frac{m}{4(2m-1)}-(t-\frac{1}{2})^2}dt.$$

The antiderivative of $\sqrt{\frac{m}{4(2m-1)}-(t-\frac{1}{2})^2}$ is

$$\frac{1}{8}\left((2t-1)\sqrt{\frac{m}{2m-1}-(1-2t)^2} - \frac{m}{2m-1}\left(\tan^{-1}\left(\frac{1-2t}{\sqrt{\frac{m}{2m-1}-(1-2t)^2}}\right)\right)\right)+C.$$

Evaluating the definite integral, we have

$$\frac{1}{8}\left((\beta-1)\sqrt{\frac{m}{2m-1}-(1-\beta)^2} - \frac{m}{2m-1}\left(\tan^{-1}\left(\frac{1-\beta}{\sqrt{\frac{m}{2m-1}-(1-\beta)^2}}\right)\right)\right)-$$
$$\frac{1}{8}\left((\alpha-1)\sqrt{\frac{m}{2m-1}-(1-\alpha)^2} - \frac{m}{2m-1}\left(\tan^{-1}\left(\frac{1-\alpha}{\sqrt{\frac{m}{2m-1}-(1-\alpha)^2}}\right)\right)\right)=$$
$$\tilde{i}(\beta,m)-\tilde{i}(\alpha,m).$$

Multiplying by $m\sqrt{2m-1}$, we find that the integral equals

$$m\sqrt{2m-1}(\tilde{i}(\beta,m)-\tilde{i}(\alpha,m)),$$

as desired. □

**Theorem 4.3.** *There are at most*

$$1+m-\frac{1}{m}+m\sqrt{2m-1}(\tilde{i}(\lfloor\frac{m}{2}\rfloor+1,m)-\tilde{i}(2,m)+\tilde{i}(m,m)-\tilde{i}(\lfloor\frac{m}{2}\rfloor,m))$$

*solutions for probabilities greater than or equal to* $\frac{m}{2m-1}$ *for* $m\in\mathbb{N}$.

*Proof.* First, because $-\lceil x \rceil \leq -x$, we have

$$1 + \sum_{i=2}^{m} (i + 1 - \lceil \tilde{x}(m, y) \rceil) \leq 1 + \sum_{i=2}^{m} (i + 1 - \tilde{x}(m, y)) . \tag{12}$$

We will now simplify this expression. One can verify that $i + 1 - \frac{m-2i+2mi}{2m} = \frac{m+2y}{2m}$. Thus, our expression is

$$\sum_{i=2}^{m} \frac{m + 2i}{2m} + \frac{m\sqrt{1 + 8i + \frac{4i^2}{m^2} - \frac{4i(1+2i)}{m}}}{2m}.$$

Note that $\sum_{i=2}^{m} \frac{m+2i}{2m} = m - \frac{1}{m}$. We wish to bound the rest of the sum with an integral. Because $\frac{m\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2m}$ defines a downwards facing parabola centered at $\frac{m}{2}$, we must split the sum into two parts–one from 2 to $\lfloor \frac{m}{2} \rfloor$, which is monotically increasing, and one from $\lfloor \frac{m}{2} \rfloor + 1$ to $m$, which is monotically decreasing. We have

$$\begin{aligned}
&\sum_{i=2}^{m} \frac{m\sqrt{1 + 8i + \frac{4i^2}{m^2} - \frac{4i(1+2i)}{m}}}{2m} = \\
&\sum_{i=2}^{\lfloor \frac{m}{2} \rfloor} \frac{m\sqrt{1 + 8i + \frac{4i^2}{m^2} - \frac{4i(1+2i)}{m}}}{2m} + \sum_{\lfloor \frac{m}{2} \rfloor + 1}^{m} \frac{m\sqrt{1 + 8i + \frac{4i^2}{m^2} - \frac{4i(1+2i)}{m}}}{2m} \leq \\
&\int_{i=2}^{\lfloor \frac{m}{2} \rfloor + 1} \frac{m\sqrt{1 + 8i + \frac{4i^2}{m^2} - \frac{4i(1+2i)}{m}}}{2m} + \int_{\lfloor \frac{m}{2} \rfloor}^{m} \frac{m\sqrt{1 + 8i + \frac{4i^2}{m^2} - \frac{4i(1+2i)}{m}}}{2m} = \\
&m\sqrt{2m-1}(\tilde{i}(\lfloor \frac{m}{2} \rfloor + 1, m) - \tilde{i}(2, m) + \tilde{i}(m, m) - \tilde{i}(\lfloor \frac{m}{2} \rfloor, m))
\end{aligned}$$

where the last equality is by Lemma 4.2. Combining, we find an upper bound of

$$1 + m - \frac{1}{m} + m\sqrt{2m-1}(\tilde{i}(\lfloor \frac{m}{2} \rfloor + 1, m) - \tilde{i}(2, m) + \tilde{i}(m, m) - \tilde{i}(\lfloor \frac{m}{2} \rfloor, m)),$$

as desired. □

**Remark.** *The upper bound for the number of solutions for probabilities greater than or equal to* $\frac{m}{2m-1}$ *is* $\mathcal{O}(m^{\frac{3}{2}})$.

*Proof.* Note that $1 + m - \frac{1}{m}$ is $\mathcal{O}(m)$, so it is $\mathcal{O}(m^{\frac{3}{2}})$. Since $\tan^{-1}(x)$ is $\mathcal{O}(1)$ asymptotically, $\tilde{i}(c, m)$ has the same growth rate as $(c-1)\sqrt{\frac{m}{2m-1} - (1-c)^2}$. Thus,

$$m\sqrt{2m-1}(\tilde{i}(\lfloor \frac{m}{2} \rfloor + 1, m) - \tilde{i}(2, m) + \tilde{i}(m, m) - \tilde{i}(\lfloor \frac{m}{2} \rfloor, m))$$

is $\mathcal{O}(m^{\frac{3}{2}})$. As such, the number of solutions is $\mathcal{O}(m^{\frac{3}{2}})$. □

**Theorem 4.4.** *There is a lower bound of*

$$1+\frac{m-1}{m}+m\sqrt{2m-1}(\tilde{i}(\lfloor\frac{m}{2}\rfloor,m)-\tilde{i}(1,m)+\tilde{i}(m+1,m)-\tilde{i}(\lfloor\frac{m}{2}+1\rfloor,m))$$

*solutions for probabilities greater than or equal to* $\frac{m}{2m-1}$.

*Proof.* We note that, because $-x-1<-\lceil x\rceil$, we have

$$1+\sum_{i=2}^{m}\left(i+1-\lceil\tilde{x}(m,y)\rceil\right)\geq 1+\sum_{i=2}^{m}\left(i-\tilde{x}(m,y)\right). \tag{13}$$

Simplifying this expression, we have

$$1+\sum_{i=2}^{m}\left(\frac{m+2i}{2m}-1+\frac{\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2}\right).$$

We have $\sum_{i=2}^{m}\frac{m+2i}{2m}-1=\frac{m-1}{m}$. We will, again, bound the remaining terms by replacing the sum with an integral. We have

$$\sum_{i=2}^{m}\frac{m\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2m}=$$

$$\sum_{i=2}^{\lfloor\frac{m}{2}\rfloor}\frac{m\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2m}+\sum_{\lfloor\frac{m}{2}\rfloor+1}^{m}\frac{m\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2m}\geq$$

$$\int_{i=1}^{\lfloor\frac{m}{2}\rfloor}\frac{m\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2m}+\int_{\lfloor\frac{m}{2}\rfloor+1}^{m+1}\frac{m\sqrt{1+8i+\frac{4i^2}{m^2}-\frac{4i(1+2i)}{m}}}{2m}=$$

$$m\sqrt{2m-1}(\tilde{i}(\lfloor\frac{m}{2}\rfloor,m)-\tilde{i}(1,m)+\tilde{i}(m+1,m)-\tilde{i}(\lfloor\frac{m}{2}+1\rfloor,m))$$

Combining, we have a lower bound of

$$1+\frac{m-1}{m}+m\sqrt{2m-1}(\tilde{i}(\lfloor\frac{m}{2}\rfloor,m)-\tilde{i}(1,m)+\tilde{i}(m+1,m)-\tilde{i}(\lfloor\frac{m}{2}+1\rfloor,m)),$$

as desired.

□

**Remark.** *The lower bound for the number of solutions for probabilities greater than or equal to* $\frac{m}{2m-1}$ *is* $\mathcal{O}(m^{\frac{3}{2}})$.

*Proof.* Note that $1+\frac{m-1}{m}$ is $\mathcal{O}(1)$. Again, $\tilde{i}(c,m)$ has the same growth rate as $(c-1)\sqrt{\frac{m}{2m-1}-(1-c)^2}$. Thus,

$$m\sqrt{2m-1}(\tilde{i}(\lfloor\frac{m}{2}\rfloor,m)-\tilde{i}(1,m)+\tilde{i}(m+1,m)-\tilde{i}(\lfloor\frac{m}{2}+1\rfloor,m))$$

is $\mathcal{O}(m^{\frac{3}{2}})$. As such, the number of solutions is $\mathcal{O}(m^{\frac{3}{2}})$. □

## 4.2 Solutions for Probabilities Between $\frac{m}{2m-1}$ and $\frac{m+1}{2m+1}$

We will now approximate the number of solutions for probabilities between $\frac{m}{2m-1}$ (the probability associated with $(m, m)$) and $\frac{m+1}{2m+1}$ (the probability associated with $(m+1, m+1)$). Unfortunately, we were unable to find appropriate formal bounds on this quantity, but we can approximate it with our previous results:

**Remark.** *Letting $n = m + 1$, there are approximately*

$$n - \frac{1}{n} + n\sqrt{2n-1}(\tilde{i}(\lfloor \frac{n}{2} \rfloor + 1, n) - \tilde{i}(2, n) + \tilde{i}(n, n) - \tilde{i}(\lfloor \frac{n}{2} \rfloor, n)) -$$
$$m - \frac{1}{m} + m\sqrt{2m-1}(\tilde{i}(\lfloor \frac{m}{2} \rfloor + 1, m) - \tilde{i}(2, m) + \tilde{i}(m, m) - \tilde{i}(\lfloor \frac{m}{2} \rfloor, m))$$

*solutions for probabilities $P(x, y)$ with $\frac{m}{2m-1} < P(x, y) < \frac{m+1}{2m+1}$.*

*Proof.* We can approximate the amount of probabilities in the desired range by taking an approximation of the amount of probabilities greater than or equal to $\frac{m+1}{2m+1}$ and subtracting an approximation of the number of probabilities greater than or equal to $\frac{m}{2m-1}$. We do this with the upper bound found in Theorem 4.3.

□

Note that, because the above approximation is essentially the slope of a function that was $\mathcal{O}(m^{\frac{3}{2}})$, it is $\mathcal{O}(m^{\frac{1}{2}})$.

Additionally, a search was conducted to find the number of probabilities between $\frac{m}{2m-1}$ and $\frac{m+1}{2m+1}$. Fitting parabola to the first three calculated maxima and to the first three minima results in the following two conjectures, respectively: (see Figure 2):

**Conjecture 3.** *The upper bound is $\sqrt{2m+2} - 3$, which is met infinitely often.*

**Conjecture 4.** *The lower bound is $\sqrt{\frac{m}{2} + 1} - 2$.*

The actual value of the number of probabilities coinciding with Conjecture 3 seem to have some relation with the families given in Theorem 3.8. For example, observe that $m = 2k^2 - 1$ iff $2m + 2 = (2k)^2$ and that $\frac{m}{2m-1}$ has a solution in Family 1 iff $m = 2k^2$ for $k \geq 2$. As such, $\sqrt{2m+2} \in \mathbb{Z}$ iff $m + 1$ has a solution in Family 1. It seems that the presence of this family increases the number of probabilities between $\frac{m}{2m-1}$ and $\frac{m+1}{2m+1}$, resulting in the true value meeting the conjectured value.

# References


[1] George E. Andrews. *Number theory.* Dover Publications, Inc., New York, 1994. Corrected reprint of the 1971 original [Dover, New York; MR0309838 (46 #8943)].

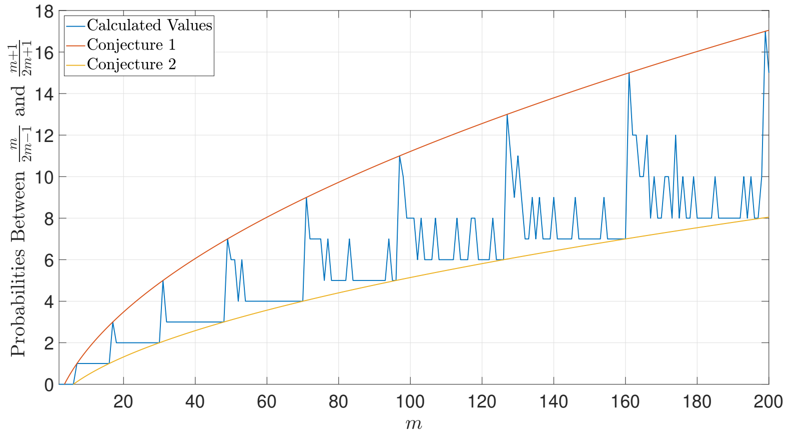


Figure 2: Actual Number of Probabilities Between $\frac{m}{2m-1}$ and $\frac{m+1}{2m+1}$ vs Conjectured Upper and Lower Bounds

[2] Kieran Hilmer, Ron Lycan, and Vadim Ponomarenko. Odds inversion problem with replacement. *To Appear in American Mathematical Monthly.*

[3] Nabil Lasmar, Cécile Mailler, and Olfa Selmi. Multiple drawing multi-colour urns by stochastic approximation. *J. Appl. Probab.*, 55(1):254–281, 2018.

[4] Robert K. Moniot. Solution of an odds inversion problem. *The American Mathematical Monthly*, 128(2):140–149, 2021.

[5] National Museum of Mathematics in New York City. *Varsity Math*, Week 117, 2017.

[6] Erol Peköz, Adrian Röllin, and Nathan Ross. Pólya urns with immigration at random times. *Bernoulli*, 25(1):189–220, 2019.

[7] Aguech Rafik, Lasmar Nabil, and Selmi Olfa. A generalized urn with multiple drawing and random addition. *Ann. Inst. Statist. Math.*, 71(2):389–408, 2019.

[8] Aguech Rafik and Selmi Olfa. Unbalanced multi-drawing urn with random addition matrix. *Arab J. Math. Sci.*, 26(1-2):57–74, 2020.