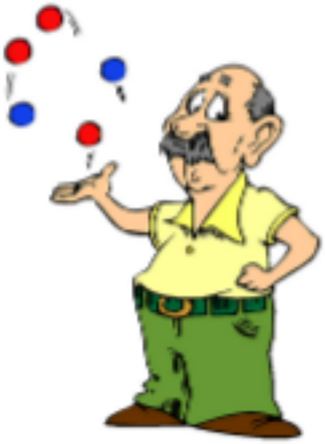
Beyond Varsity Math:
The red-and-blue-balls puzzle

An exploration of Diophantine equations

Robert K. Moniot
Fordham University



Juggler image modified from https://commons.wikimedia.org/wiki/File:Juggling_man.jpg by Hhamilton / Creative Commons Attribution-Share Alike.

Introduction

"From a bag containing red and blue balls, two are removed at random. The chances are 50-50 that they will differ in color. What were the possible numbers of balls initially in the bag?"

This is the National Museum of Mathematics' Varsity Math puzzle 117.

We will solve that problem and then explore the solutions for other values of the probability.

- The results of this exploration are in an article accepted for publication in the *American Mathematical Monthly*, “Solution of an Odds Inversion Problem,” by Robert K. Moniot, 2020 (to appear).

Solution of *Varsity Math* problem

Let *r*, *b* be the numbers of red and blue balls, respectively. Probability of picking red, then blue:

$$P(\text{red}, \text{blue}) == \frac{r}{b+r} \cdot \frac{b}{b+r-1}$$

Probability of picking blue, then red:

$$P(\text{blue}, \text{red}) == \frac{b}{b+r} \cdot \frac{r}{b+r-1}$$

These are equal. So, probability of picking balls of different colors are:

$$P(\text{red}, \text{blue}) + P(\text{blue}, \text{red}) == \frac{2\,b\,r}{(b+r)\,(b+r-1)}$$

Equate this to 50% or 1/2.

$$\frac{2\,b\,r}{(b+r)\,(b+r-1)} == \frac{1}{2}$$

Cross multiply and expand:

$$4\,b\,r == (b+r)\,(b+r-1) == b^2 + 2\,b\,r + r^2 - (b+r)$$

$$b^2 - 2\,b\,r + r^2 == b + r$$

$$(b-r)^2 == b + r$$

This implies total number of balls $b + r$ must be a square. Call it n^2. Then

$$b + r == n^2, \quad b - r == \pm n$$

Taking the positive branch so $b \geq r$, and adding and subtracting these, we obtain

$$2\,b == n^2 + n \Rightarrow b == \frac{n^2+n}{2} \qquad 2\,r == n^2 - n \Rightarrow r == \frac{n^2-n}{2}$$

So *r* and *b* are successive triangular numbers, whose sum is square. That’s pretty cool!

First 10 solutions

In[∘]:=

$$\texttt{TableForm}\left[\texttt{Table}\left[\left\{\frac{n^2-n}{2},\ \frac{n^2+n}{2},\ n^2\right\},\ \{n,\ 1,\ 10\}\right],\ \texttt{TableHeadings}\to\{\texttt{None},\ \{\texttt{"r"},\ \texttt{"b"},\ \texttt{"r+b"}\}\}\right]$$

Out[∘]//TableForm=

```
r      b      r+b
0      1      1
1      3      4
3      6      9
6      10     16
10     15     25
15     21     36
21     28     49
28     36     64
36     45     81
45     55     100
```

The first row formally satisfies the equation, but is not admissible since there need to be at least 2 balls.

Generalizing the problem

That was fun. Now, what can we say about other odds than 50-50?

We seek to answer the questions:

- For a given odds, is there a solution to the problem? That is, does there exist a pair of numbers of red and blue balls that produce a given probability of drawing balls of different colors?
- If there is a solution, is the number of solutions finite or infinite?
- For instances with a finite number of solutions, is there a way to list them all?
- For instances with an infinite number of solutions, can we obtain a formula or recurrence that will generate as many solutions as desired?
- Are the methods used able to find all solutions? Or do they miss some?

All of these questions can be answered very satisfactorily.

Some examples

For the following probability ratios there is **no solution.**

$$\frac{4}{9}\quad \frac{5}{6}\quad \frac{3}{4}\quad \frac{3}{8}\quad \frac{98\,058}{176\,501}$$

For these probabilities there are **only these solutions:**

$$\frac{8}{15} \to \begin{matrix} 2 & 4 \\ 4 & 6 \\ 7 & 8 \\ 8 & 8 \end{matrix} \qquad \frac{12}{25} \to 9 \;\; 16$$

For the following probabilities there are an **infinite** number of solutions:

Odds	Smallest Solutions	
2 / 7	1	6
	6	30
	30	145
	145	696
	696	3336
6 / 13	4	9
	9	18
	18	34
	70	126
	126	225
4 / 51	281 063 265 339 959 257	6 593 004 860 639 355 312
	6 593 004 860 639 355 312	154 654 550 959 684 890 576
	154 654 550 959 684 890 576	3 627 788 942 691 955 573 225

Notice that a number from one solution often appears in the next.

Diophantine equation

Let the given probability be p/q, in lowest terms. We are solving

$$\frac{2\,b\,r}{(b+r)\,(b+r-1)} == \frac{p}{q}$$

Cross multiplying and rearranging, we obtain

$$p\,b^2 - 2\,(q-p)\,b\,r + p\,r^2 - p\,b - p\,r == 0 \tag{1}$$

This is a ***Diophantine equation,*** i.e. an equation in integers.

We will call any pair of integers (r, b) satisfying Equation (1) a ***formal*** solution.

Not all formal solutions are ***admissible*** as solutions to the original problem. For a solution to be admissible it must satisfy

$r \geq 0, \;\; b \geq 0$

(we cannot have a negative number of balls) and

$r + b \geq 2$

(we need at least two balls in the bag in order to be able to draw two out).

Trivial solutions

Equation (1) has three ***trivial*** solutions, i.e. formal solutions that hold regardless of the values of p and q:

$(r, b) == (0, 1),\ (0, 0),\ (1, 0)$

These are not admissible solutions, but they will prove useful in finding admissible solutions.
When plugged into the original odds equation,

$$\frac{2\,b\,r}{(b+r)\,(b+r-1)} == \frac{p}{q}$$

they yield $p/q == 0/0$. This is undefined, which is how they are able to satisfy the equation for any values of p and q.

Symmetry

Equation (1) is ***symmetric*** in r and b:

If (r, b) is a solution then (b, r) is also.

To list just ***distinct*** solutions, we impose the condition $r \le b$. (We already did that in solving the *Varsity Math* problem by only using the positive sign on n.)

Reverse search

To get a feel for the problem, I wrote a Python program to generate the odds for all values of red, blue less than 1000.

Pseudocode:

```
for b in {1..max}
  for r in {1..b}
    odds=(2 r b)/((r+b)(r+b-1))
    append {odds,{r,b}} to result
sort result by odds
```

Python has a `Fraction` class and unlimited-size integers, which make this easy.

The results showed interesting patterns, such as the reappearance of a number from one solution in the next noted earlier.

The recycling recurrence

Often solutions for a particular p/q occur in series in which the same number appears in two solutions, once as the smaller value and once as the larger, i.e.

$(r, b),\ (b, x)$

where $r < b < x$. (The *Varsity Math* case shows this pattern, repeating indefinitely.)

By equating the probability ratios for these two solutions and imposing the condition *r* ≠ *x* one obtains this formula:

In[•]:=
$$\texttt{Solve}\left[\left\{\frac{2\ r\ b}{(r+b)\ (r+b-1)} == \frac{2\ b\ x}{(b+x)\ (b+x-1)},\ x \neq r\right\},\ x\right]$$

Out[•]=
$$\left\{\left\{x \to \frac{(-1+b)\ b}{r}\right\}\right\}$$

I call this the ***recycling recurrence*** since the new solution re-uses one of the values in the old solution.

If *r* < *b* then the formula produces a larger solution, whereas if *r* ≥ *b* it goes the other direction.

The result *x* given by the formula is not necessarily integer. One can show that if *p* == 1 or *p* == 2, the recurrence produces integers ad infinitum. For *p* > 2, it goes fractional after a few steps, with only 2 or 3 integer solutions in any series.

Recycling recurrence examples

Here is an example of a series of solutions for the probability of drawing balls of different color equal to 2/7.

```
In[•]:= TableForm[RecurrenceTable[{r[i + 1] == b[i],
           b[i + 1] == b[i] (b[i] - 1) / r[i], r[1] == 1, b[1] == 6}, {r[i], b[i]}, {i, 6}],
         TableHeadings → {None, {"r", "b"}}]
```

```
Out[•]//TableForm=
r        b
1        6
6        30
30       145
145      696
696      3336
3336     15 985
```

The recycling pattern continues indefinitely.

Here is an example of a series of solutions for which only three are integer pairs. The probability ratio is 10/21:

```
In[•]:= TableForm[RecurrenceTable[{r[i + 1] == b[i],
           b[i + 1] == b[i] (b[i] - 1) / r[i], r[1] == 2 / 5, b[1] == 2}, {r[i], b[i]}, {i, 5}],
         TableHeadings → {None, {"r", "b"}}]
```

```
Out[•]//TableForm=
r      b
2/5    2
2      5
5      10
10     18
18     153/5
```

Change of variables

Equation (1) simplifies considerably if we change variables. Let

$$t == b + r,\quad v == b - r$$

Plug in, and after some algebra, Equation (1) becomes

$$(q - 2\,p)\,t^2 + 2\,p\,t - q\,v^2 == 0 \tag{2}$$

Much simpler. But we can do even better.

Complete the square on *t*.

To make the next steps clearer, set $a == (q - 2\,p)$:

$$a\,t^2 + 2\,p\,t - q\,v^2 == 0$$

First step is to make the coefficient of t^2 square by multiplying by *a*:

$$a^2\,t^2 + 2\,p\,a\,t - q\,a\,v^2 == 0$$

Add p^2 to both sides:

$$a^2\,t^2 + 2\,p\,a\,t + p^2 - q\,a\,v^2 == p^2$$
$$(a\,t + p)^2 - q\,a\,v^2 == p^2$$

Put back $a == (q - 2\,p)$ and let

$$u == a\,t + p == (q - 2\,p)\,t + p$$

Then the equation becomes

$$u^2 - q\,(q - 2\,p)\,v^2 == p^2 \tag{3}$$

We put Equations (1), (2) and (3) into formulas for convenient use later, with *p* and *q* as parameters.

```
In[∘]:= rbequation[r_, b_] := p b^2 - 2 (q - p) b r + p r^2 - p b - p r == 0

In[∘]:= tvequation[t_, v_] := (q - 2 p) t^2 + 2 p t - q v^2 == 0

In[∘]:= uvequation[u_, v_] := u^2 - q (q - 2 p) v^2 == p^2
```

The discriminant *D*

Set $D == q(q - 2\,p)$, and $f == p^2$. Then Equation (3) is

$$u^2 - D\,v^2 == f \tag{4}$$

Call *D* the ***discriminant*** because it determines the character of the equation.

- $D < 0 \Rightarrow$ **ellipse**
- $D > 0 \Rightarrow$ **hyperbola**

- If $D == 0$ Equation (4) breaks down. Recall that in deriving Equation (3) from Equation (2) (the t, v equation), we multiplied by $a == (q - 2\,p)$ to complete the square, so (3) is not valid when that is 0. Equation (2) is a **parabola** for that case. It was already solved as the *Varsity Math* problem, so no problem.

Character of equation vs *p*/*q*

D == q (q − 2 p)

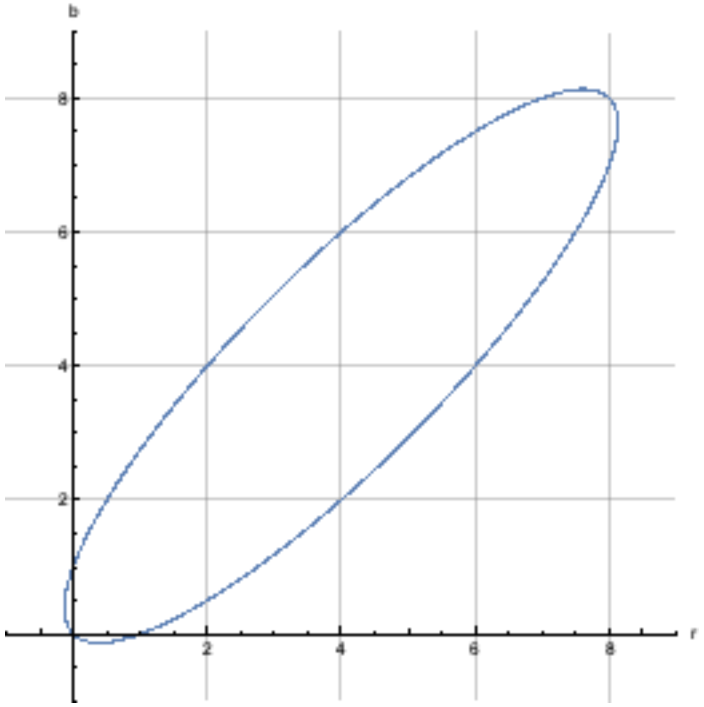
- Elliptic: $D < 0 \Rightarrow q < 2\,p \Rightarrow \frac{p}{q} > \frac{1}{2}$
- Parabolic: $D == 0 \Rightarrow q == 2\,p \Rightarrow \frac{p}{q} == \frac{1}{2}$
- Hyperbolic: $D > 0 \Rightarrow q > 2\,p \Rightarrow \frac{p}{q} < \frac{1}{2}$

Elliptical case: *D* < 0

For the elliptic case I have not found any methods other than direct search, i.e. testing all possibilities. Since the ellipse is a finite curve, direct search can in principle find all solutions that exist.

It is helpful to graph an example. Here is one, $p/q == 8/15$, that has several integer solutions. First we graph it in r-b space. The grid lines show where solutions occur, for instance (4, 6).

```
Plot[b /. Solve[rbequation[r, b] /. {p → 8, q → 15}], {r, -1, 9},
 PlotRange → {{-1, 9}, {-1, 9}}, AspectRatio → 1,
 GridLines → {{2, 4, 6, 8}, {2, 4, 6, 8}}, AxesLabel → {"r", "b"}]
```

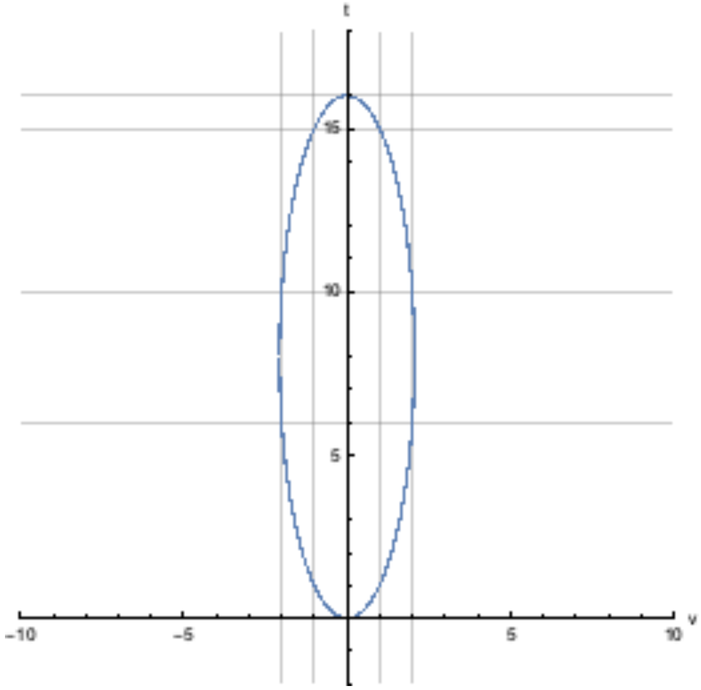


Elliptical case, example plotted in *t*,*v*-space

Now plot the $p / q == 8 / 15$ example in *t*-*v* space. This ellipse is aligned with the axes and is tangent to the horizontal axis.

```
In[∘]:= Plot[t /. Solve[tvequation[t, v] /. {p → 8, q → 15}], {v, -5 / 2, 5 / 2},
         PlotRange → {{-10, 10}, {-2, 18}}, AspectRatio → 1,
         GridLines → {{-2, -1, 1, 2}, {6, 10, 15, 16}}, AxesLabel → {"v", "t"}]
```

Out[∘]=

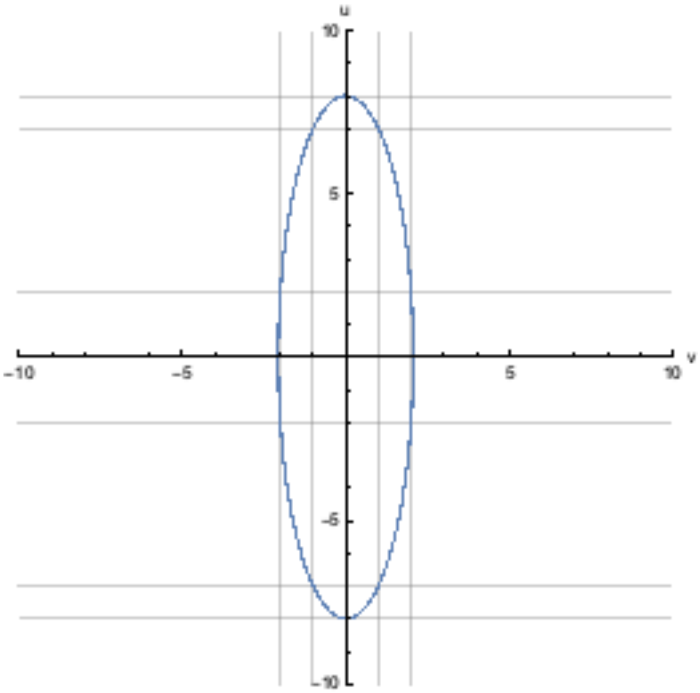


Elliptical case, example plotted in *u*-*v* space

Now plot the $p / q == 8 / 15$ example in *u*-*v* space. This ellipse is aligned with the axes and centered on the origin.

In[∘]:=
```
Plot[u /. Solve[uvequation[u, v] /. {p → 8, q → 15}], {v, -5 / 2, 5 / 2},
 PlotRange → {{-10, 10}, {-10, 10}}, AspectRatio → 1,
 GridLines → {{-2, -1, 1, 2}, {-8, -7, -2, 2, 7, 8}}, AxesLabel → {"v", "u"}]
```

Out[∘]=



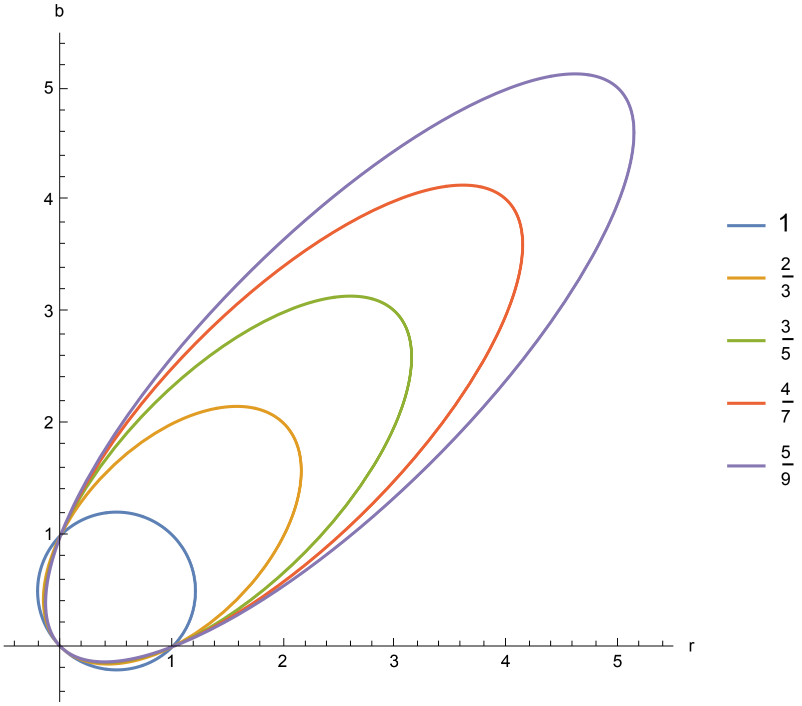
Elliptical cases vs. *p*/*q*

Plot of ellipse in *r*-*b* space for a series of odds ratios of form *p*/(2 *p* – 1):

In[∘]:=

```
listforplot = Table[{b /. Solve[rbequation[r, b] /. q → 2 p - 1, b]}, {p, 5}];
Plot[listforplot,
 {r, -1 / 2, 11 / 2},
 PlotRange → {{-1 / 2, 11 / 2}, {-1 / 2, 11 / 2}}, AspectRatio → 1,
 PlotLegends → Table[p / (2 p - 1), {p, 5}], AxesLabel → {"r", "b"}]
```

Out[∘]=

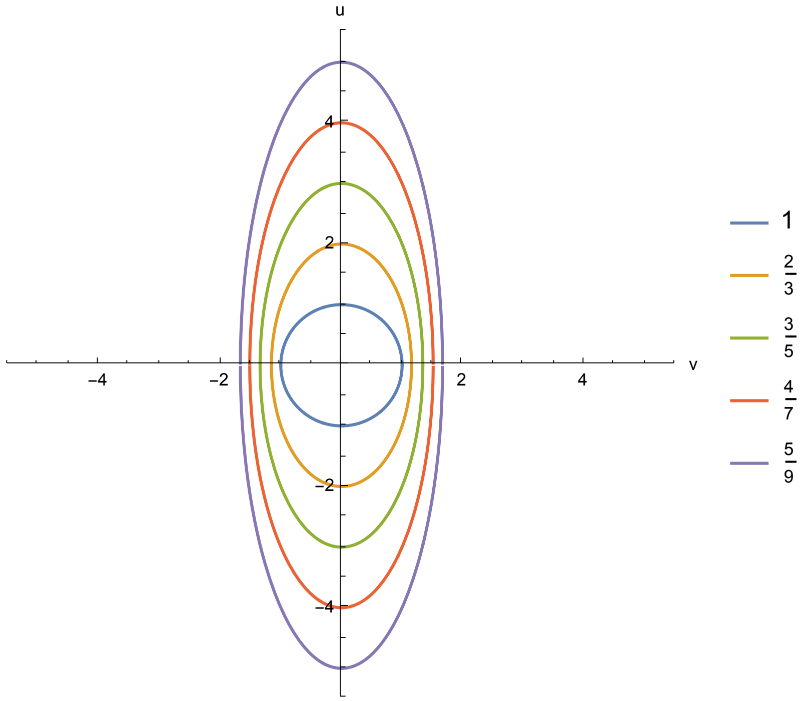


For $p/q == 1$, it is a circle. As $p/q \to 1/2$, the ellipses elongate, approaching a parabola at $p/q == 1/2$.

Elliptical cases in *u*-*v* space

Same set of cases plotted in *u*-*v* space.

```
listforplot = Table[u /. Solve[uvequation[u, v] /. q → 2 p - 1, u], {p, 5}];
Plot[listforplot,
 {v, -2, 2},
 PlotRange → {{-11 / 2, 11 / 2}, {-11 / 2, 11 / 2}}, AspectRatio → 1,
 PlotLegends → Table[p / (2 p - 1), {p, 5}], AxesLabel → {"v", "u"}]
```



The widths grow slowly. This means a search on *v* is more efficient than a search on *u* or on *r* or *b*.

Symmetry allows the testing to be limited to positive *u*, *v* values (first quadrant arc). Need to use both signs on *u* to get all solutions for *t*. Can use only positive *v* to keep *r* ≤ *b*.

Elliptical case: range of *r*, *b*

$$p\,b^2 - 2\,(q - p)\,b\,r + p\,r^2 - p\,b - p\,r == 0 \qquad (1,\ \text{repeated})$$

Solve for the far endpoint where *b* == *r*:

$$p\,r^2 - 2\,(q - p)\,r^2 + p\,r^2 - p\,r - p\,r == 0$$

$$4\,p\,r^2 - 2\,q\,r^2 - 2\,p\,r == 0$$

$$r\,((2\,p - q)\,r - p) == 0$$

$$r_{\max} == \frac{p}{2\,p - q} == \frac{p / q}{2\,p / q - 1}$$

Solve for *p*/*q* value for a given $r_{\max}$:

$$\frac{p}{q} == \frac{r_{\max}}{2\,r_{\max} - 1}$$

For $r_{\max} == 5$, $p/q == 5/9$. For larger p/q ratios, solutions must have r, $b < 5$. We can exhaustively list them.

- Note: if $q == 2p - 1$, then $r_{\max} == p$. This implies there are solutions for all probabilities of form $p/(2p-1)$. The recycling recurrence from (p, p) gives neighboring solutions $(p-1, p)$ and $(p, p-1)$. These three solutions at the far vertex of the ellipse are the symmetrical counterparts of the three trivial solutions around the origin.

Elliptical case: exhaustive enumeration

Table enumerating all distinct solutions for r, $b \leq 5$:

In[∘]:=
```
TableForm[Table[{{r, b, 2 r b / ((r + b) (r + b - 1))}}, {b, 1, 5}, {r, 1, b}]]
```

Out[∘]//TableForm=

$$\begin{array}{lllll}
1\ 1\ 1 & & & & \\
1\ 2\ \frac{2}{3} & 2\ 2\ \frac{2}{3} & & & \\
1\ 3\ \frac{1}{2} & 2\ 3\ \frac{3}{5} & 3\ 3\ \frac{3}{5} & & \\
1\ 4\ \frac{2}{5} & 2\ 4\ \frac{8}{15} & 3\ 4\ \frac{4}{7} & 4\ 4\ \frac{4}{7} & \\
1\ 5\ \frac{1}{3} & 2\ 5\ \frac{10}{21} & 3\ 5\ \frac{15}{28} & 4\ 5\ \frac{5}{9} & 5\ 5\ \frac{5}{9}
\end{array}$$

Some of these ratios are below the threshold $p/q == 5/9$ so the search is not exhaustive for them.

For ratios in the range 5/9 to 1 the list is exhaustive. These ratios are:

$$\frac{5}{9},\ \frac{4}{7},\ \frac{3}{5},\ \frac{2}{3},\ \frac{1}{1}$$

All of their solutions appear in this table. No other ratios in this range have any solutions.

For example: 3/4, 4/5, 5/6, 5/7, 6/7, 5/8, 7/8 all have no solution.

This covers nearly half the range of possible probabilities!

Elliptical case: direct search in *u*-*v* space

Maximum v is for $u == 0$:

In[∘]:=
```
Solve[uvequation[u, v] /. u → 0, v]
```

Out[∘]=

$$\left\{\left\{v \to -\frac{p}{\sqrt{2\,p\,q - q^2}}\right\}, \left\{v \to \frac{p}{\sqrt{2\,p\,q - q^2}}\right\}\right\}$$

We choose the positive root,

$$v_{\max} == \frac{p}{\sqrt{q\,(2\,p - q)}} == \frac{p\,/\,q}{\sqrt{2\,p\,/\,q - 1}}$$

For an ellipse that extends to a million red balls,

$$\frac{p}{q} == \frac{r_{\max}}{2\, r_{\max} - 1}, \quad r_{\max} == 10^{6} \Rightarrow \quad \frac{p}{q} == \frac{10^{6}}{2 \times 10^{6} - 1}$$

(This is very close to 1/2.) Plugging this into the formula for $v_{\max}$ yields

In[∘]:= $\mathtt{Floor}\left[\frac{p}{\sqrt{q\,(2\,p - q)}} \;/.\; \{p \to 10^{6},\, q \to 2 \times 10^{6} - 1\}\right]$

Out[∘]= 707

Searching this range for solutions takes 0.04 seconds on my laptop. The only solutions found are for the far endpoint (b, b) and adjacent (b – 1, b):

r	b
999 999	1 000 000
1 000 000	1 000 000

- Even quite large solutions are within reach of direct search.

Hyperbolic case: $D > 0$

The hyperbolic case can be divided into two categories:

- *D* square: solution by factoring the equation
- *D* non-square: solution by continued fractions

Hyperbolic case, *D* square

If *D* is square, then Equation (4) can be factored:

$$u^{2} - D\, v^{2} == f \to \left(u - \sqrt{D}\; v\right)\left(u + \sqrt{D}\; v\right) == f$$

Here $\sqrt{D}$ is integer by assumption. The two terms on the left must correspond to divisors of f.

Let divisors of f be $\{d_1, d_2, \ldots, d_k\}$. Set

$$\left(u - \sqrt{D}\; v\right) == d_i, \quad \left(u + \sqrt{D}\; v\right) == \frac{f}{d_i}$$

Solution:

$$u == \frac{f + d_i^{2}}{2\, d_i}, \quad v == \frac{f - d_i^{2}}{2\, \sqrt{D}\; d_i}$$

Reject any solutions that are not integer.

Exchanging d_i and f/d_i simply changes the sign of v, so only the divisors $d_i \leq \sqrt{f}$ need to be tested.

Negative divisors simply change sign of u, v, so only positive divisors need to be tested.

At most one solution (u, v) for each divisor, so total number of solutions is finite. Some ratios have no solutions.

Hyperbolic case, *D* square example: *p*/*q* == 12/25

Here is an example that has an admissible solution.

$$D == q\,(q-2\,p) == 25\,(25-24) == 25 == 5^2$$
$$f == p^2 == 12^2 == 144$$

Divisors of *f*: {1, 2, 3, 4, 6, 8, 9, 12, 16, 18, 24, 36, 48, 72, 144}

Too many for a demo. Use only the one that gives a solution.

$$d == 2 \Rightarrow u == \frac{f+d^2}{2\,d} == \frac{144+4}{2\times 2} == \frac{148}{4} == 37, \quad v == \frac{f-d^2}{2\,\sqrt{D}\,d} == \frac{144-4}{2\times 5\times 2} == \frac{140}{20} == 7$$

$$t == \frac{u-p}{q-2\,p} == \frac{37-12}{25-24} == 25$$

$$r == \frac{t-v}{2} == \frac{25-7}{2} == \frac{18}{2} == 9, \quad b == \frac{t+v}{2} == \frac{25+7}{2} == \frac{32}{2} == 16$$

Hyperbolic case, *D* square example: *p*/*q* == 4/9

$$D == q\,(q-2\,p) == 9\,(9-8) == 9 == 3^2$$
$$f == p^2 == 4^2 == 16$$

Divisors of *f*: {1, 2, 4, 8, 16}.

Trying them all, one finds none lead to an admissible solution.
This case has no solutions.

Hyperbolic case, *D* nonsquare

$$u^2 - D\,v^2 == f == p^2 \qquad (4,\ \text{repeated})$$

Divide through by p^2:

$$\left(\frac{u}{p}\right)^2 - D\left(\frac{v}{p}\right)^2 == 1, \text{ or } x^2 - D\,y^2 == 1$$

This is known as the ***Pell equation.*** It has been well studied for a few centuries.

When $D > 0$ is nonsquare, its solutions are found from x/y == a **convergent** of the **continued fraction** $\sqrt{D}$.

These concepts are explained in the next few sections.

Continued fractions

Continued fraction of a real number *r*:

$$r == a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots}}}$$

where the a_i are integers. Continued fractions are usually written in the compact form

$$r == [a_0, a_1, a_2, \ldots]$$

There is a simple algorithm to compute continued fractions.

If *r* is rational, the fraction *terminates*. Otherwise it *continues indefinitely.*

If *r* is the irrational root of a quadratic equation with rational coefficients, the continued fraction *repeats.*

Convergents

Terminating the continued fraction at any point yields a rational number.

a_0

$$a_0 + \frac{1}{a_1} == \frac{a_0 a_1 + 1}{a_1}$$

$$a_0 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{a_2 (a_1 a_0 + 1) + a_0}{a_2 a_1 + 1}$$

...

These are called ***convergents*** of the continued fraction for *r*.

There is a simple algorithm to generate the sequence of convergents.

The convergents are the ***best rational approximations*** to *r* using a denominator of that size. Example: 22/7 for π.

Pell equation

$$x^2 - D y^2 == 1$$

For $D > 0$ nonsquare, the Pell equation always has a nontrivial solution. (Trivial solution is $x == \pm 1$, $y == 0$.)

The smallest nontrivial solution is given by setting x/y equal to the convergent of $\sqrt{D}$ at the end of the first repeat cycle if the repeat cycle is even length; otherwise the end of the second cycle.

- Example: $p/q == 4/11$, $D == 11 (11 - 2 \times 4) == 33$

```
In[•]:= Convergents[√33]
```

$$\text{Out[•]=} \left\{5, 6, \frac{17}{3}, \frac{23}{4}, \sqrt{33}\right\}$$

Mathematica indicates the start of the repeat cycle by repeating the argument. Show that this solution works:

In[∘]:= $23^2 - 33\times 4^2$

Out[∘]= 1

Solution:

$$x == \frac{u}{p} == 23,\quad y == \frac{v}{p} == 4$$

Get *u*, *v*:

$$u == p\,x == 4\times 23 == 92,\quad v == p\,y == 4\times 4 == 16$$

Map *u* back to *t* == *r* + *b*.

$$t == \frac{u-p}{q-2\,p} == \frac{92-4}{11-2\times 4} == \frac{88}{3}$$

Fractional, not admissible. But don’t lose hope.

Formula to generate more solutions

If (x_1, y_1) is the smallest solution of the Pell equation, other solutions can be generated by

$$x_n + y_n \sqrt{D} == \left(x_1 + y_1 \sqrt{D}\right)^n,\quad n == 1,\ 2,\ \ldots \tag{5}$$

When the expression on the RHS is expanded, after collecting terms it is always in the form of $x + y\sqrt{D}$.

- Example: $p/q == 4/11$, $D == 33$, $(x_1, y_1) == (23, 4)$

$$\left(23 + 4\sqrt{33}\right)^2 == 23^2 + 2\times 4\times 23\sqrt{33} + 16\times 33 == 1057 + 184\sqrt{33} \Rightarrow x_2 == 1057,\quad y_2 == 184$$

Verify that this is also a solution to the Pell equation.

In[∘]:= $1057^2 - 33\times 184^2$

Out[∘]= 1

Convert back to *r*, *b*.

$$u == p\,x == 4\times 1057 == 4228,\quad v == p\,y == 4\times 184 == 736$$

$$t == \frac{u-p}{q-2\,p} == \frac{4228-4}{11-2\times 4} == \frac{4224}{3} == 1408$$

$$r == \frac{t-v}{2} == 336,\quad b == \frac{t+v}{2} == 1072$$

Infinite number of solutions

Example shows smallest solution of Pell equation does not necessarily give admissible (*r*, *b*) solution.

However, it can be proved that if (x_1, y_1) is the smallest nontrivial solution of the Pell equation, then the solution given by

$$\left(x_1 + y_1 \sqrt{D}\right)^{2n}, \quad n > 0$$

always yields an admissible (*r*, *b*) solution.

Hence the number of solutions to our problem is always infinite for *D* > 0 and non-square.

Recurrence starting from trivial solutions

The Pell recurrence can be unwound algebraically to work directly on *r*, *b*.

It can then be applied to the trivial solutions (0, 0), (0, 1), and (1, 0) to generate solutions.

Starting from (0, 0) gives the Pell solution series. Starting from (0, 1) or (1, 0) gives neighbors related via the recycling recurrence.

We're not done yet!

We have a method that always finds solutions to the hyperbolic, non-square *D* case.

However, it misses some solutions. Example: $p/q == 20/41$. Solving the Pell equation yields

r	b
17 280	23 680

Additional solutions found using the recurrence are only larger. Yet the following are solutions:

r	b
16	25
85	120
120	168
552	760

Note that the first is not related to another via the recycling recurrence. The next two are a doublet, not a triplet.
The Pell equation method can only yield solutions that are members of a triplet related by the recycling recurrence.
This is rooted in the fact that its solutions are the same as found by applying the Pell recurrence to the trivial solutions, which are three in number.

It can be shown that if *p* is a prime greater than 2, then the method based on the Pell equation is in fact complete. But for composite *p*, there can be additional solutions.

Finding all solutions

Hua (1982) gives a method that finds all solutions to Equation (4).

$$u^2 - D v^2 == f \qquad (4,\ \text{repeated})$$

If $|f| < \sqrt{D}$, then if it has a solution, the solution will be found among the convergents of $\sqrt{D}$, similarly to solving the Pell equation, but with the difference that the solution is not the last convergent in the cycle. One has to search the convergents for solutions, and there may be more than one in the first cycle.

If this inequality is not satisfied, then we proceed as follows.

Hua, Luogeng (Loo Keng) (1982). *Introduction to Number Theory,* translated from the Chinese by Peter Shiu. Springer-Verlag. Chapter 11.

Method of solution

The method requires seeking integers *l*, *h* satisfying

$$l^2 == D + f\,h$$

which can be solved by searching. It suffices to search using the range $-h_{\max} < h < h_{\max}$ where

$$h_{\max} == \mathrm{Max}\left(\left|\frac{f}{4}\right|, \left|\frac{D}{f}\right|\right)$$

Since $f \geq \sqrt{D}$ it is guaranteed that $h_{\max} < |f|$. Thus the RHS is always reduced. Then solve

$$u^2 - D\,v^2 == h$$

If $|h| < \sqrt{D}$ solve directly using continued fractions; otherwise repeat recursively. Once one has a solution

$$x^2 - D\,y^2 == h$$

then solutions to $u^2 - D\,v^2 == f$ are given by

$$u == \frac{D\,y \pm l\,x}{h}, \quad v == \frac{x \pm l\,y}{h}$$

One more important element is needed to ensure completeness. The continued fraction method only yields solutions in which gcd(u, v) == 1. If gcd(u, v) == $g > 1$, then g^2 must divide $f == p^2$, i.e. g must divide p. To find those solutions, then, one can divide Equation (4) by each of the divisors of p. Thus, for each g that is a divisor of p, solve

$$\left(\frac{u}{g}\right)^2 - D\left(\frac{v}{g}\right)^2 == \left(\frac{p}{g}\right)^2$$

Then multiply the solution by g to obtain (non coprime) u, v satisfying Equation (4).

This method finds all solutions.