Mathematical Spectrum

A magazine for students and teachers of mathematics
in schools, colleges and universities,
and for everyone interested in mathematics

Volume 46 2013/2014 Number 1

- Napier's Candles
- The Mathematics of Origami
- UK Personal Income
- Is the IPL 2012 Tournament Fair?

Mathematical Spectrum is a magazine for students and teachers in schools, colleges and universities, as well as the general reader interested in mathematics. It is published by the Applied Probability Trust, a non-profit-making organisation established in 1963 with the support of the London Mathematical Society. The object of the Trust is the encouragement of study and research in the mathematical sciences.

One volume of *Mathematical Spectrum* is published in each British academic year and consists of three issues, which appear in September, January and May.

Articles published in *Mathematical Spectrum* deal with the entire range of mathematical disciplines (pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, biomathematics). Both expository and historical material may be included, as well as elementary research and information on educational opportunities and careers in mathematics. There are also sections devoted to problems, to mathematics in the classroom and to computing.

From the Editor

Sophie's Diary

Coincidentally, a book arrived for review on the day Nicola Adams won the first gold medal ever awarded at an Olympic games for women's boxing, which previously had been thought to be the sole preserve of the male of the species. Marie-Sophie Germain was born on 1 April 1776 in Paris and showed an early interest in Mathematics, to the chagrin of her mother who, when Sophie announced that she wanted to be a mathematician, described her intention as 'this ridiculous pursuit unbecoming to a young lady'. Thereafter, her mother did all that she could to kill this interest, so that Sophie was reduced to studying at night in an unheated bedroom by the light of a candle.

This volume, which describes itself as a mathematical novel, presents what the young Sophie might have written in her diary. It describes how she struggled to get hold of the writings of the great mathematicians, from the ancient Greeks to those of her own day, and to understand them without the help of teachers. Through it all, her obsession was undaunted. Not for her the support that our present-day Olympic athletes enjoy. But she shared their single-minded determination to master her subject, or at least understand as much as she could. As if prejudice against women were not enough, the diary vividly describes how she lived in the thick of the events of the French Revolution. Her house was close by La Conciergerie, the prison known as 'the antechamber to the guillotine', and she and her family were in not inconsiderable danger.

After the fictional reconstruction of the diary, there is a biographical sketch of Sophie's life and an author's note describing, amongst other things, how she wrote to Lagrange using the assumed name Monsieur Le Blanc and did likewise to Gauss. When it was discovered that Monsieur Le Blanc was in fact a Mademoiselle, surprise gave way to a degree of recognition, so much so that she won the Prix de Mathématiques of the Institut de France for her work on the theory of vibrations and elastic surfaces. She made an important contribution to the eventual solution to Fermat's Last Theorem in a result now known as Germain's Theorem.

I write this on the day that the annual scramble to secure places at UK universities takes place as the Advanced level examination results are announced. Spare a thought for Sophie Germain, who was barred from attending an institute of higher education because of her sex, who was self-taught, but who was surely worthy of an Olympic medal in Mathematics!

Reference

1 D. Musielak, *Sohpie's Diary* (The Mathematical Association of America, Washington, D.C., 2012).

Napier's Candles

PAUL GLAISTER and ELIZABETH GLAISTER

A model is proposed which determines the all-important placement of divisions on Advent candles of various shapes.

Each year, as winter approaches and our thoughts turn to Christmas, we keep our eyes open for an Advent candle in readiness for lighting on December 1st. Most years the shape of the candles have been cylindrical, and figure 1 shows a vertical cross-section of one such candle. The divisions are obviously equally spaced, so that the volume of wax between them is constant, and the candle will burn for the same amount of time each day. (Our children much preferred the candle when it became quite short as that signified Christmas day was getting close!)

In the last few years, however, the candles available have been a little more varied in shape and are typically tapered, being narrower at the top than the bottom. It was only relatively

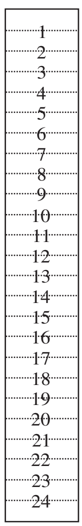


Figure 1

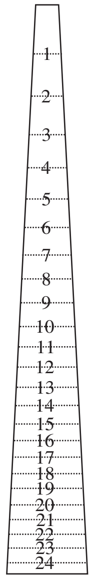


Figure 2

recently that we noticed that some of these had unequally spaced divisions, no doubt to ensure that the candle still burned for the same amount of time each day. (There were also some where the divisions remained equally spaced, which required us to be extra vigilant initially; otherwise, two days worth of wax would disappear before we had noticed, and our children's wait appeared to be shorter than it actually was!) Two particular candles which spring to mind were as follows.

The first was in the shape of an elongated, truncated pyramid with a square base, a vertical cross-section of which is shown in figure 2. The second was in the shape of an elongated, right-circular, truncated cone, whose cross-section was identical to that of the pyramid shown in figure 2. The divisions were *not* equally spaced in either case, and we wondered how these were determined.

To solve this problem, we shall assume that the rate at which the candle burns is a constant, i.e. the reduction in volume occurs at a constant rate k. (While this rate may *decrease* slightly as the width of the candle increases—because the flame is further from the wax at the edge of the candle—in practice, for an elongated candle of the shape we are considering, we shall assume that this effect is negligible.)

We first assume that the candle is in the shape of either a square-based, truncated pyramid, or a right-circular, truncated cone, whose vertical cross-section is shown in figure 3, where we have added suitable axes. Denote by $h(t)$ the height of the candle at time t days after the candle is lit (remembering that it is blown out each day once the next division is reached), and denote the corresponding width of the candle by $w(t)$. If the initial height is $h(0) = h_0$, the half-width of the top is $w(0) = w_0$, and the half-width of the base is W, then

$$w(t) = W - \frac{W - w_0}{h_0} h(t), \tag{1}$$

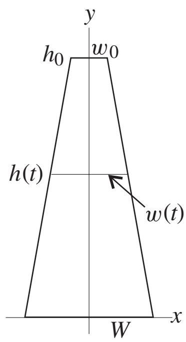


Figure 3

from the straight line relationship

$$\frac{y}{h_0} = \frac{W - x}{W - w_0} \tag{2}$$

for points (x, y) on the edge of the candle in figure 3. Furthermore, the cross-sectional area A of the candle at height $h(t)$ and half-width $w(t)$ is given by

$$A(h(t)) = \alpha w(t)^2, \tag{3}$$

where the constant $\alpha = 4$ in the case of the pyramid and $\alpha = \pi$ for the cone. Combining (1) and (3) we have

$$A(h(t)) = \alpha\left(W - \frac{W - w_0}{h_0} h(t)\right)^2, \tag{4}$$

giving $A(0) = \alpha W^2$ and $A(h_0) = \alpha w_0^2$. Now, if we denote by $V(h(t))$ the volume of the *remaining* part of the candle after it has burnt for t days and when the height is $h(t)$, then

$$V(h(t)) = V(h_0) - kt, \tag{5}$$

where k is the constant rate of burning and $V(h_0)$ is the initial volume of the candle. Given that the candle is completely burned through after 25 days, i.e. when $t = 25$, then $V(h(25)) = 0$; thus. from (5) we have $k = V(h_0)/25$, and, hence,

$$V(h(t)) = V(h_0)\left(1 - \frac{t}{25}\right). \tag{6}$$

Using the standard result for the volume of a pyramid or a cone as $\frac{1}{3}$ *base area* $\times$ *height*, and noting that the height of the candle shown in figure 3 if it was extended to the vertex would be $h_0 W/(W - w_0)$ (by setting $x = 0$ in (2)), then the volume of the candle remaining when the height is $h(t)$ is

$$V(h(t)) = \frac{1}{3} A(0) \frac{h_0 W}{W - w_0} - \frac{1}{3} A(h(t)) \left(\frac{h_0 W}{W - w_0} - h(t)\right). \tag{7}$$

This is obtained by subtracting the volume extending upward from $y = h(t)$ to the vertex from the volume extending upward from $y = 0$ to the vertex. Substituting (4) into (7) and simplifying gives

$$V(h(t)) = \frac{1}{3}\frac{\alpha h_0}{W - w_0}\left[W^3 - \left(W - \frac{W - w_0}{h_0}h(t)\right)^3\right]. \tag{8}$$

Furthermore, with $h(0) = h_0$ initially, then (8) gives

$$V(h_0) = \frac{1}{3}\frac{\alpha h_0}{W - w_0}[W^3 - w_0^3]. \tag{9}$$

Finally, if we substitute (8) and (9) into (6), and rearrange to write $h(t)$ explicitly in terms of t, then

$$h(t) = h_0 \frac{W - [w_0^3 + (W^3 - w_0^3)t/25]^{1/3}}{W - w_0},$$

giving the height after t days. Hence, the values $h(1), h(2), \ldots, h(24)$ give the heights at which the divisions should be made, and we can see that these are independent of α, i.e. the divisions are identical for both the truncated pyramid and the truncated cone, provided the height and widths at the top and bottom are the same. Figure 2 shows the divisions for a candle where $w_0 = 0.75$ cm, $W = 1.5$ cm, and $h_0 = 30$ cm.

Last year the candle we chose had a *curved* shape similar to that shown in figure 4. The divisions were also unequally spaced, and again we wondered how these were determined. The shape reminded us of a 'log' graph, reflected in the x-axis and rotated about the y-axis to generate a *volume of revolution*. Equation (6) still holds, assuming a constant rate of burning. However, referring to figure 5, this time the volume of the candle remaining after t days can

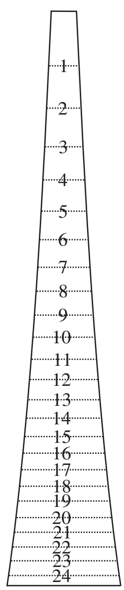


Figure 4

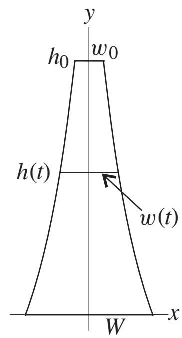


Figure 5

be determined by integration as

$$V(h(t)) = \pi \int_0^{h(t)} x^2 \,\mathrm{d}y = \pi \int_0^{h(t)} f(y)^2 \,\mathrm{d}y, \tag{10}$$

where $x = f(y)$ represents the equation of the curved edge of the candle. Furthermore, from (10), the initial volume of the candle is

$$V(h_0) = V(h(0)) = \pi \int_0^{h_0} f(y)^2 \,\mathrm{d}y, \tag{11}$$

where the initial half-width w_0, half-width of the base W, and initial height h_0, satisfy $w_0 = f(h_0)$ and $W = f(0)$. (Note that $x = f(y) = W - (W - w_0)y/h_0$ in the case of a conically shaped candle, and substituting this expression into (10) and (11) gives the results in (8) and (9) where $\alpha = \pi$.)

Suppose that we model the edge of the candle with the curve in the shape of a natural logarithm reflected in the x-axis as

$$\frac{y}{h_0} = \frac{-\ln(x/W)}{\ln(W/w_0)} = \frac{\ln(x/W)}{\ln(w_0/W)},$$

i.e.

$$x = W \exp\left\{-\frac{y}{h_0}\ln\left(\frac{W}{w_0}\right)\right\} = W\left(\frac{w_0}{W}\right)^{y/h_0}, \tag{12}$$

which satisfies $w_0 = f(h_0)$ and $W = f(0)$. Substituting (12) into (10), and integrating, gives

$$V(h(t)) = \pi W^2 \int_0^{h(t)} \left(\frac{w_0}{W}\right)^{2y/h_0} \mathrm{d}y = \frac{\pi h_0 W^2[(w_0/W)^{2h(t)/h_0} - 1]}{\ln(w_0/W)^2}, \tag{13}$$

and, using (13), the initial volume is

$$V(h_0) = V(h(0)) = \frac{\pi h_0 W^2[(w_0/W)^2 - 1]}{\ln(w_0/W)^2}. \tag{14}$$

Finally, if we substitute (13) and (14) into (6), and rearrange to write $h(t)$ explicitly in terms of t, then

$$h(t) = h_0 \frac{\ln\{(w_0/W)^2 - [(w_0/W)^2 - 1]t/25\}}{\ln(w_0/W)^2}, \tag{15}$$

giving the height after t days. Hence, the values $h(1), h(2), \ldots, h(24)$ give the heights at which the divisions should be made. Figure 4 shows the divisions for a candle where $w_0 = 0.75$ cm, $W = 1.5$ cm, and $h_0 = 30$ cm.

And then the penny dropped...

The variable spacing here reminded us of that day some years ago (over 40 in the case of one of us!) when we were introduced to the *slide rule*. The logarithmic scale we see in figure 4, and based on (15), is merely a reincarnation of the slide rule!

To see this more clearly, consider a candle which is designed to burn for 9 days *instead* of 25, and denote by $d(t)$ the distance between divisions from the top of the candle before it is first lit. Then

$$d(t) = h_0 - h(t) = h_0 - h_0 \frac{\ln\{(w_0/W)^2 - [(w_0/W)^2 - 1]t/9\}}{\ln(w_0/W)^2}, \tag{16}$$

using a modification of (15). After some simplification, (16) can be rewritten as

$$d(t) = h_0 \frac{\ln\{1 + [(W/w_0)^2 - 1]t/9\}}{\ln(W/w_0)^2} = h_0 \frac{\log_{10}\{1 + [(W/w_0)^2 - 1]t/9\}}{\log_{10}(W/w_0)^2}, \tag{17}$$

where we have converted from natural logarithms to base-10 logarithms. With $W = 1.5$ cm, $w_0 = W/\sqrt{10} = 1.5/\sqrt{10} \approx 0.47$ cm, and $h_0 = 30$ cm, then equation (17) becomes $d(t) = 30\log_{10}(1 + t)$, and we replicate precisely, as shown in figure 6, a standard slide rule where the markings start with 1 at the top through to 10 at the bottom.

Who knows, the logarithmic candle may have been a favourite of the Scottish mathematician John Napier who discovered logarithms!

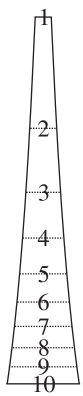


Figure 6

Paul Glaister *lectures in mathematics at Reading University. His research interests include computational fluid dynamics, numerical analysis, and perturbation methods, as well as mathematics and science education.*

Elizabeth Glaister *teaches mathematics at Kendrick School in Reading, as well as lecturing in mathematics at Reading University. While they sleep a little better than they used to as Christmas approaches, our children (18 and 21!) still maintain a keen interest in the demise of the Advent candle as it burns through each year.*

Nomo Triples

JONNY GRIFFITHS

This article introduces a new triplet of natural numbers called a Nomo triple that arises from a problem in classical mechanics concerning a stationary mass.

Dr Hugh Hunt of Cambridge University presented this lovely question at Maths Jam 2010:

> Imagine the system in figure 1 released from rest with integer masses x, y, and z (in kilograms), where $x < y$. Pulleys are smooth and light, strings are light and inextensible. If z remains stationary, what values are possible for (x, y, z)?

Let T be the tension in the string attached to z (in newtons), let t be the tension in the string connecting x and y (in newtons), and let a be the acceleration of x and y (in m per s^2). Then $yg - t = ya$ and $t - xg = xa$, so

$$a = g\frac{y-x}{x+y} \quad \text{and} \quad t = xg + xg\frac{y-x}{x+y}.$$

We also have $T = 2t$ and $T = zg$, so

$$z = \frac{2t}{g} = 2x\left(1 + \frac{y-x}{x+y}\right) = \frac{4xy}{x+y}.$$

Note that T and so t and a remain constant over time. Our solutions, if you like, define a new integer triple, $(x, y, 4xy/(x+y))$ with $x < y$. I have called this a Nomo triple (hereafter called an NT), since it arises from our 'no motion' question. The ordering here is that x must be the smallest element of the triple, since

$$\frac{4xy}{x+y} - x = \frac{3xy - x^2}{x+y} > 0,$$

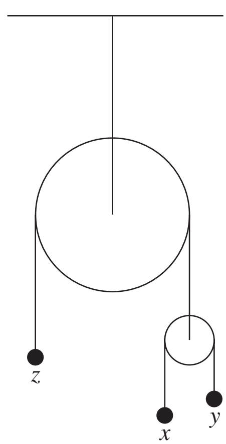


Figure 1 The 'no motion' problem.

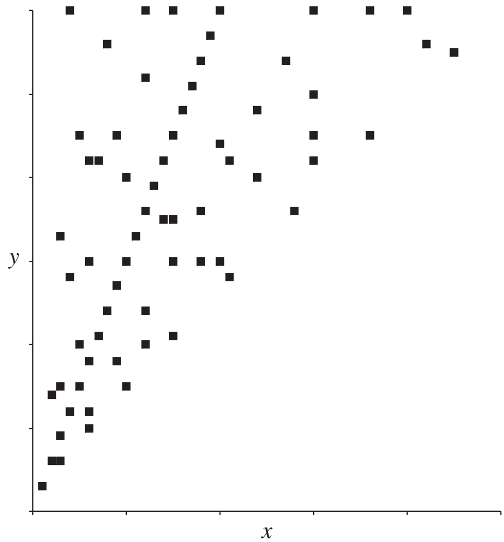


Figure 2 Early NTs.

but $y < 4xy/(x+y)$ and $y > 4xy/(x+y)$ are both possible, since $(3, 6, 8)$ and $(2, 14, 7)$ are both NTs (there is also the special case $(k, 3k, 3k)$ where $y = 4xy/(x+y)$).

How many NTs are there? It is easy to search with a computer, and the answer is that they are not rare. The early ones are shown in figure 2. Note that

$$\begin{aligned}(x, y, z) \text{ is an NT} \quad &\iff \quad z = \frac{4xy}{x+y} \\ &\iff \quad zk = \frac{4xyk^2}{kx+ky} \\ &\iff \quad (kx, ky, kz) \text{ is an NT,}\end{aligned}$$

giving the NTs lying on straight lines through the origin in figure 2. We can thus define a *primitive* NT (x, y, z) to be one where $\gcd(x, y, z) = 1$. The first few primitive NTs are given in table 1, ordered by their sum.

Can we find a parametrisation for primitive NTs, akin to that for primitive Pythagorean triples? Finding one that yields all primitive NTs and only primitive NTs would seem to ask a lot, but one where we account for all primitive NTs, with as few as possible nonprimitive NTs included, seems much more possible.

Using Microsoft EXCEL® to chart the first few primitive NTs yields figure 3. Careful pattern-spotting yields the following two parametrisations.

Type 1: $(\frac{1}{2}n(n+2m-1), \frac{1}{2}(n+2m-1)(n+4m-2), n(n+4m-2))$.

Type 2: $((2a-1)(2a+2b-1), (2a+2b-1)(2a+4b-1), 2(2a+4b-1)(2a-1))$.

Here $n, m, a,$ and b are positive integers. It is easily checked that these both always give NTs. In figure 3 the type-1 family is shown by filled squares and the type-2 family by open squares.

Table 1 Early NTs ordered by their sum.

x	y	z	$x+y+z$
1	3	3	7
3	6	8	17
2	14	7	23
3	15	10	28
6	10	15	31
5	20	16	41
3	33	11	47
10	15	24	49
5	45	18	68
15	21	35	71

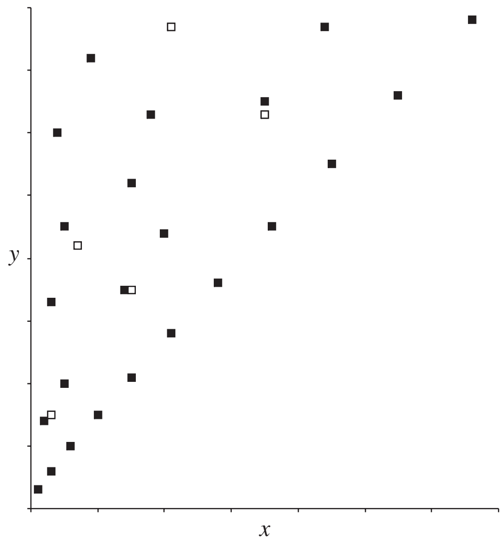


Figure 3 Early primitive NTs.

Putting $x = \frac{1}{2}n(n+2m-1)$, finding m, and then substituting m into $y = \frac{1}{2}(n+2m-1)(n+4m-2)$ gives

$$y = \frac{4}{n^2}x^2 - x.$$

While putting $x = (2a-1)(2a+2b-1)$, finding b, and then substituting b into $y = (2a+2b-1)(2a+4b-1)$ gives

$$y = \frac{2}{(2a-1)^2}x^2 - x.$$

We thus have two families of parabolas on which the primitive NTs lie. In figure 4 the type-1 parabola families are shown as solid lines, the type-2 families are shown as dashed lines, while

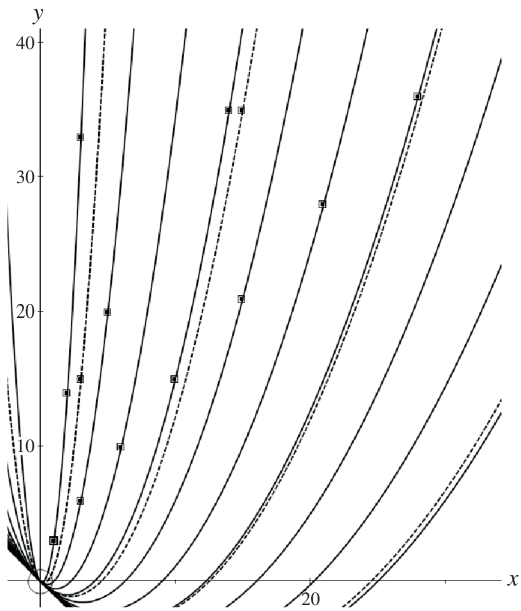


Figure 4 Type-1 and type-2 parabolas.

the squares represent the small primitive NTs (which are points on the parabolas with natural number coordinates).

I conjecture that all primitive NTs are included in these parametrisations (along with some that are not primitive). A computer check for primitive NTs where x and y are less than 5000 supports this hypothesis. The set of type-1 NTs and the set of type-2 NTs are disjoint, since

$$\frac{2}{(2a-1)^2} = \frac{4}{n^2}$$

yields $n = -\sqrt{2}(2a-1)$ or $n = \sqrt{2}(2a-1)$, which is impossible.

Jonny Griffiths *teaches at Paston College in Norfolk, where he has been for the last twenty years. He has also taught at St Dominic's Sixth Form College in Harrow-on-the-Hill. He has studied maths at Cambridge University, with the Open University, and at the University of East Anglia. Possible claims to fame include being an ex-member of Harvey and the Wallbangers, a popular band in the 1980s, and playing the character Stringfellow on the childrens' television programme Playdays.*

The Mathematics of Origami

SUDHARAKA PALAMAKUMBURA

Origami is a widespread art form gaining popularity among mathematicians for its remarkable ability to perform geometric constructions. This article provides a brief introduction to the mathematical aspects of origami and shows how origami can be used to perform two well-known constructions, angle trisection and doubling the cube, which are impossible to solve using only the compass and the straight edge.

Introduction

Origami is the art of creating sculptures through paper folding. The word 'origami' is derived from two Japanese words: 'Oru', which means 'fold', and 'Kami', which means 'paper'. As the name implies, in origami the folding of paper is used to create designs. Cutting and gluing papers are not considered origami techniques. It is believed that origami started around 105 AD with the invention of paper. Although the exact place where origami was born is subject to much debate, it is believed to have begun in Korea, China, or Japan (see reference 1).

Origami began as a form of art and inspires both young and old due to the vast number of design possibilities that it opens (for some examples, see http://www.langorigami.com). In some countries, origami is part of the culture; in Japan newly wedded couples and newborn babies are presented with a thousand origami paper cranes to symbolize long life and happiness, while in China the burning of particular origami models is a tradition at funerals (see reference 2).

In the late nineteenth century there was a spur of interest in studying the mathematical aspects of origami. In 1893 T. Sundara Row published a book named *Geometric Exercises in Paper Folding* (reference 3) which illustrates the geometrical aspects of origami. Since then, considerable efforts have been made in order to mathematize origami. In 1985 a Japanese mathematician named Humiaki Huzita presented a set of six axioms which determine what is constructible through origami folds. Later, in 2001, Koshiro Hatori added one more axiom to the collection (see reference 1). The seven axioms are widely known as the Huzita–Hatori axioms. One of the pioneers of modern origami, Robert J. Lang, recently proved that this system of axioms is complete (reference 4), i.e. every single fold in origami corresponds to one of the seven axioms.

Huzita–Hatori axioms

It is traditional practice to use square paper in origami. Hence, in this article *paper* refers to square paper hereafter. A *line* in origami is a crease made by a paper fold or the boundary of the paper. Initially, a paper consists of four lines: its boundaries. A *point* is the intersection of two lines. The Huzita–Hatori axioms define what can be constructed via a single fold in terms of lines and points and are illustrated in figure 1.

(A1) Given two points, one can fold a crease line through them.

(A2) Given two points, one can fold a crease along their perpendicular bisector, folding one point on top of the other.

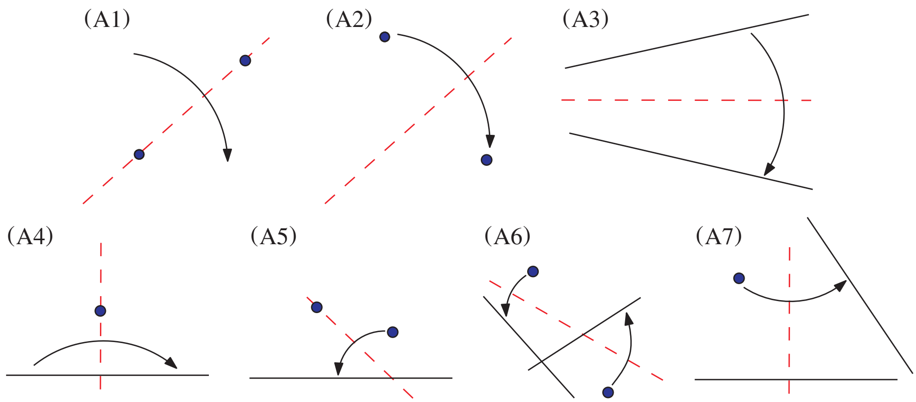


Figure 1 Huzita–Hatori axioms, dashed lines represent creases and solid lines represent existing lines.

(A3) Given two lines, one can fold their bisector crease, folding one line on top of the other.

(A4) Given a point and line, one can crease through the point perpendicular to the line, folding the line onto itself.

(A5) Given two points and a line, one can fold a crease through one point that maps the other point onto the line.

(A6) Given two points and two lines, one can fold a crease that simultaneously maps one point to one line and the other point to the other line.

(A7) Given one point and two lines, one can fold a crease perpendicular to one line so that the point maps to the other line.

The power of origami

One of the interesting aspects of origami is its ability to go beyond traditional compass and straightedge (unmarked ruler) constructions. Angle trisection and doubling the cube are two major problems which haunted mathematicians for over 2000 years. In 1837 the French mathematician Pierre Laurent Wantzel proved that both of these constructions are impossible using only the straightedge and compass (reference 5). However, both constructions can be carried out using origami folds.

Trisecting an angle

Trisecting an angle is the problem of dividing a given angle into three equal parts using the compass and straightedge. In 1980 Hishashi Abe discovered an elegant method for trisecting a given angle using origami folds (reference 1). Figure 2 illustrates Abe's trisection method. The given angle is marked in the figure as θ. Note that all the folds in this construction agree with the Huzita–Hatori axioms.

1. The first fold (figure 2(a)) finds the bisector crease ab corresponding to the two boundaries of the paper. This is in accordance with (A3).

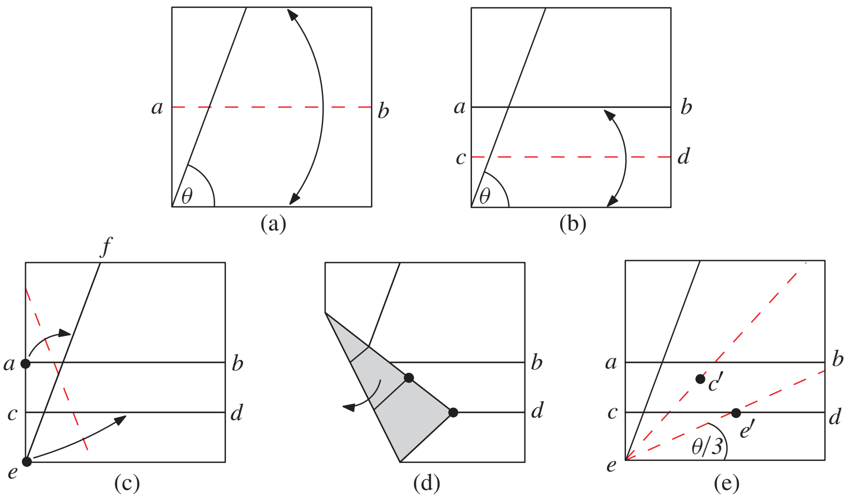


Figure 2 Abe's method to trisect an angle. Dashed lines represent creases and solid lines represent existing lines.

2. Similarly, the second fold (figure 2(b)) finds the bisector crease cd corresponding to the lower boundary and ab. This is also in accordance with (A3).

3. The third fold (figure 2(c) and (d)) is by (A6) and folds e onto cd and a onto ef; c maps to c'.

4. The last two folds (figure 2(e)) are by (A1) and find the lines ec' and ee'.

To see how Abe's method trisects an angle, we need to consider the crease pattern of the method (figure 3). The crease hg maps the three points $a, c,$ and e to $a', c',$ and e', respectively. Therefore,

$$\triangle hei \equiv \triangle he'i \quad \Longrightarrow \quad \angle aee' = \angle a'e'e, \tag{1}$$

$$ae = a'e', \tag{2}$$

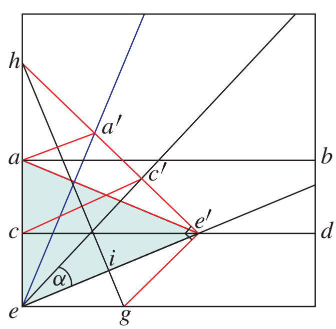


Figure 3 Crease pattern of Abe's construction for trisecting an angle.

and

$$ee' \text{ is common to both } \triangle aee' \text{ and } \triangle a'ee'. \tag{3}$$

By (1), (2), and (3),

$$\triangle aee' \equiv \triangle a'e'e. \tag{4}$$

Since ce' is the perpendicular bisector of ae, then $ae' = ee'$ and, from (4), $ea' = ee'$. Since c is the midpoint of ae and c' is the midpoint of $a'e'$, it follows that

$$\angle a'ec' = \angle e'ec' = \alpha. \tag{5}$$

Note that

$$ec' \perp a'e' \quad \Longrightarrow \quad \angle ee'c' = 90^\circ - \alpha \quad \Longrightarrow \quad \angle ee'g = \alpha.$$

Since $eg = ge'$, $\angle e'eg = \angle ee'g = \alpha$. Therefore, by (5), $\angle a'ec' = \angle e'ec' = \angle e'eg = \alpha$ and the lines ea' and ee' trisect the angle θ.

Doubling the cube

Doubling the cube is an ancient problem that was proved unsolvable using the compass and straightedge. Given a cube of volume V and side length a, the problem is whether a cube with volume $2V$ and side length $a\sqrt[3]{2}$ can be constructed. The problem can be reduced to dividing a given line segment ab into two parts, ac and cb, such that $ac : cb = 1 : \sqrt[3]{2}$. In 1985, Peter Messer found a method to double the cube through origami (reference 1; see figure 4).

1. The first fold (figure 4(a)) finds the bisector crease ab. This is in accordance with (A3).
2. The point a' is brought onto b (figure 4(b) and (c)). This is an (A2) fold. Here we find the new point c which is a trisection point of the side pq. The remaining trisection point d can be obtained by making the bisector crease between pc. Hence, the square is divided into three equal parts (figure 4(d)).

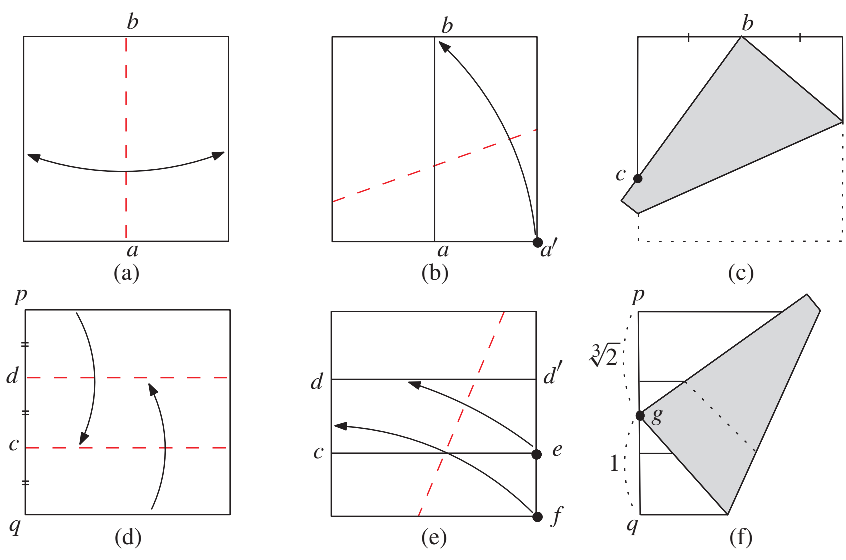


Figure 4 Messer's method of doubling the cube.

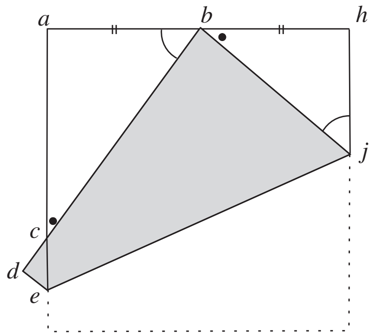


Figure 5 Trisecting the side of the square.

3. Finally, the points e and f are mapped onto the two lines dd' and pq, respectively (figure 4(e) and (f)). This fold is by (A6). The new point g is found, which divides the side pq by the ratio, $gq\colon pg = 1\colon \sqrt[3]{2}$.

In the first half of this construction, the paper is divided into three equal parts, as illustrated in figure 5. Let l be the length of one side of the square. Then $ab = bh = l/2$. Consider $\triangle bhj$. By the Pythagorean relation $bh^2 + hj^2 = jb^2$, we have $(l/2)^2 + hj^2 = jb^2$. However, $hj + jb = l$; therefore,

$$\left(\frac{l}{2}\right)^2 + (l - jb)^2 = jb^2 \quad \Longrightarrow \quad jb = \frac{5l}{8}, \qquad hj = \frac{3l}{8}.$$

Hence, the ratio between the sides of $\triangle bhj$ is, $hj\colon bh\colon jb = 3\colon 4\colon 5$ ($\triangle bhj$ turns out to be a scalar multiple of a $3\colon 4\colon 5$ right-angled triangle!). Since $\triangle bhj$ and $\triangle abc$ are similar triangles,

$$\frac{ac}{ab} = \frac{bh}{hj} \quad \Longrightarrow \quad \frac{ac}{l/2} = \frac{4}{3} \quad \Longrightarrow \quad ac = \frac{2l}{3}.$$

Hence, c is a trisection point. After dividing the paper into three equal parts the remainder of the construction (see figure 4(e) and (f)) divides pq in the ratio $1\colon \sqrt[3]{2}$. Figure 6 depicts the underlying geometry. Let $pg = y$ and $gq = x$. Considering $\triangle gqr$, we have

$$\begin{aligned}
& l = x\tan\theta + x\sec\theta \\
\Longrightarrow \quad & l = x(\tan\theta + \sqrt{1+\tan^2\theta}) \\
\Longrightarrow \quad & l = x\left(\frac{\sin\theta}{\sqrt{1-\sin^2\theta}} + \sqrt{1 + \frac{\sin^2\theta}{1-\sin^2\theta}}\right) \\
\Longrightarrow \quad & l = x\sqrt{\frac{1+\sin\theta}{1-\sin\theta}}.
\end{aligned} \tag{6}$$

Considering $\triangle gst$, we have

$$\sin\theta = \frac{y - l/3}{l/3}.$$

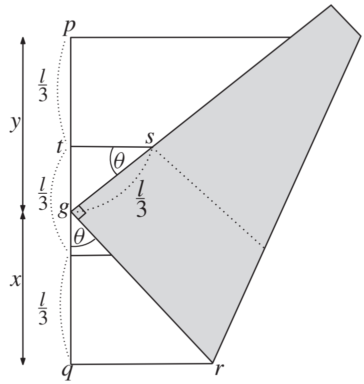


Figure 6 Dividing the side by the ratio $1 : \sqrt[3]{2}$.

Substituting this result into (6) yields

$$l = x\sqrt{\frac{3y}{2l - 3y}}\,.$$

Since $l = x + y$,

$$x + y = x\sqrt{\frac{3y}{2(x + y) - 3y}}\,.$$

Simplification of this expression leads to $(x/y)^3 = \frac{1}{2}$; therefore,

$$\frac{qg}{gp} = \frac{1}{\sqrt[3]{2}}.$$

Acknowledgments The author would like to thank and dedicate this article to all the members of the amazing mathematical forum *Math Help Boards* (http://www.mathhelpboards.com/forum/). Their contributions made me look at mathematics from a different perspective.

References

1 E. D. Demaine and J. O'Rourke, *Geometric Folding Algorithms: Linkages, Origami, Polyhedra* (Cambridge University Press, 2007).

2 F. Temko, *Paper Pandas and Jumping Frogs* (China Books, San Francisco, 1998).

3 T. S. Row, *Geometric Exercises in Paper Folding* (The Open Court Publishing Company, Chicago, 1917).

4 R. J. Lang, Origami and geometric constructions, http://www.langorigami.com/science/math/hja/hja.php (2010).

5 D. M. Burton, *The History of Mathematics: An Introduction* (McGraw–Hill, New York, 2006).

Sudharaka Palamakumbura *is an undergraduate student at the University of Peradeniya, Sri Lanka. His mathematical interests include cryptography and coding theory. In his spare time, he likes to ride his bicycle and listen to music.*

Simson's Ellipse

G. T. VICKERS

The wonderful result of Simson states that the feet of the three perpendiculars from any point on the circumcircle of a triangle onto the three sides of that triangle are always collinear. While this is very clean and precise, it is unfortunate that the restriction (that the chosen point lie on the circumcircle) is so restrictive. This article removes this requirement. Naturally, any new result must be of a somewhat different nature and is necessarily a little less 'clean'. Use is made of some of the ideas presented in reference 1, particularly regarding the notion of an envelope.

1. Introduction

Let the point P lie on the circumcircle of the triangle ABC, and let O be the centre of this circle. Let D, E, F be the feet of the perpendiculars from P onto the lines BC, CA, AB, respectively. It is a classical result that the points D, E, F lie on a line, usually called the *Simson's line* of P. Reference 1 was concerned with how this line varies as P moves around the circle. It was shown that there is a curve (referred to as *Simson's envelope*) which touches every Simson's line. Put another way, every Simson's line is tangential to the Simson's envelope. It is also true that if P does not lie on the circumcircle, then the points D, E, F are not collinear. But one might expect that these three points do have some interesting properties. Indeed, it is shown later that if P moves around a circle concentric with the circumcircle, then the area of the triangle DEF is constant. This prompted an investigation into other possible properties of D, E, F when P is restricted to such a circle.

It is convenient (as in reference 1) to take the radius of the circumcircle to be unity so that, with axes centred on O, the coordinates of the points A, B, C are $(\cos\theta_i, \sin\theta_i)$ for $i = 1, 2, 3$. Also, the axes are presumed to have been chosen so that

$$\theta_1 + \theta_2 + \theta_3 = 0. \tag{1}$$

If P is a point on the circumcircle of the triangle ABC, then its coordinates may be taken to be $(\cos\phi, \sin\phi)$. The points D, E, F (defined as above) form the Simson's line of P. A subsidiary result, which is important here, is that this line passes through the midpoint of PH, where H is the orthocentre of ABC.

It is shown in reference 1 that the coordinates of the centre N of the nine-point circle of the triangle ABC are (x_0, y_0), where

$$x_0 = \tfrac{1}{2}(\cos\theta_1 + \cos\theta_2 + \cos\theta_3), \qquad y_0 = \tfrac{1}{2}(\sin\theta_1 + \sin\theta_2 + \sin\theta_3),$$

and that the coordinates of H are $(2x_0, 2y_0)$. Also, if P is allowed to move around the circumcircle, then the set of Simson's lines has an envelope (called *Simson's envelope* in reference 1) given parametrically by

$$x = x_0 + \cos\phi + \tfrac{1}{2}\cos 2\phi \quad \text{and} \quad y = y_0 + \sin\phi - \tfrac{1}{2}\sin 2\phi.$$

With

$$x - x_0 = r\cos\psi \quad \text{and} \quad y - y_0 = r\sin\psi \tag{2}$$

(i.e. using polar coordinates centred on N), the above equations give

$$1 - \cos 3\psi = \frac{(1+2r)(3-2r)^3}{64r^3}. \tag{3}$$

(The derivation of this equation is somewhat involved; the equation is given partly because it shows the threefold symmetry of the curve and partly because it is surprisingly simple. Similar results are given later and the same comments apply.) This curve will play an important role in what follows and is here given the name of the *basic (Simson's) envelope*. Reference 1 gives much more information about this envelope.

2. Simson's ellipse

Suppose now that P is an arbitrary point, not necessarily on the circumcircle. The points D, E, F may still be constructed, and a reasonable question is whether some generalisation of Simson's line can be invented which has interesting properties. In particular, can a figure be constructed for each P which gives a 'nice' envelope as P moves in a circle concentric with the circumcircle?

One possible way of enlarging the notion of a Simson's line is the following. Given any triangle ABC (with orthocentre H) and any point P, construct D, E, F as the feet of the perpendiculars from P to BC, CA, AB, respectively. Consider the ellipse which has the midpoint of PH as its centre and which passes through the points D, E, F. This (unique) ellipse will be called the *Simson's ellipse* of the point P. (I doubt that Simson would recognise this as his, but he is not likely to complain.) A typical case is shown in figure 1.

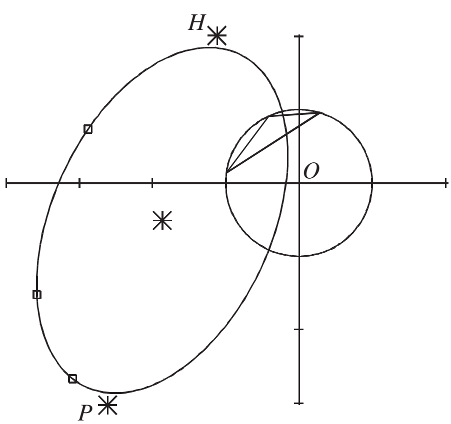


Figure 1 A triangle with its circumcircle and the Simson's ellipse for an arbitrary point P. The feet of the three perpendiculars are shown as squares and the three starred points are P, the orthocentre H, and their midpoint which is the centre of the ellipse.

Using the axes described earlier and letting $OP = R$, the coordinates of P can be written as $(R\cos\phi, R\sin\phi)$. With parameter t, the equation of this ellipse is

$$\begin{aligned} x &= x_0 + \tfrac{1}{2}[R\cos\phi - R\cos(t+\phi) - \cos t], \\ y &= y_0 + \tfrac{1}{2}[R\sin\phi + R\sin(t+\phi) - \sin t]. \end{aligned} \tag{4}$$

The values $\theta_1, \theta_2, \theta_3$ for t give the points D, E, F, respectively. The derivation of this result is given in the appendix. Although straightforward, it is not for the faint hearted.

Before describing some of the properties of this ellipse, a digression is made to explore some features of the triangle DEF which is naturally referred to as *Simson's triangle*.

3. Simson's triangle

A standard result in plane coordinate geometry is that if the coordinates of a triangle are (x_i, y_i) for $i = 1, 2, 3$, then the area of the triangle is

$$\pm\frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}.$$

Using this result, and knowing the coordinates of A, B, C and D, E, F (the latter set from (4)), allows the calculation of the areas δ and δ' of the triangles ABC and DEF. The results are

$$\begin{aligned} \delta &= \pm 2\sin\left(\frac{\theta_2-\theta_3}{2}\right)\sin\left(\frac{\theta_3-\theta_1}{2}\right)\sin\left(\frac{\theta_1-\theta_2}{2}\right), \\ \delta' &= \pm\frac{1}{2}(R^2-1)\sin\left(\frac{\theta_2-\theta_3}{2}\right)\sin\left(\frac{\theta_3-\theta_1}{2}\right)\sin\left(\frac{\theta_1-\theta_2}{2}\right), \end{aligned}$$

and so

$$4\delta' = \delta|1 - R^2|.$$

This equation is mentioned in reference 2 and some consequences of choosing P to be various particular points are explored there. For the present purpose, it is sufficient to point out that the area of the triangle DEF is 0 (i.e. the triangle collapses to a line) if and only if P lies on the circumcircle and the area of the triangle DEF is independent of ϕ, so that, as P moves around a circle (centred on O), the value of the area does not change.

4. Properties of Simson's ellipse

As shown in the appendix, the ellipse given by (4) has axes of length $R+1$ and $|R-1|$; also, its major axis makes an angle of $(-\phi/2)$ with the x-axis. Thus, as P moves around a circle concentric with the circumcircle of triangle ABC, a family of similar ellipses is generated. Figure 2 shows that the nature of the collection of ellipses depends upon whether R is greater or less than unity. When $R = 1$, P lies on the circumcircle and the ellipse collapses to a line segment of length 2 (the diameter of the circumcircle). Also, when $R = 0$, P coincides with O and the ellipse then becomes the nine-point circle. Dr C. R. Jordan has very kindly provided a simulation of these figures (see reference 3), which provides an opportunity to explore the form of the ellipses, and their envelope, for any shape of the triangle ABC.

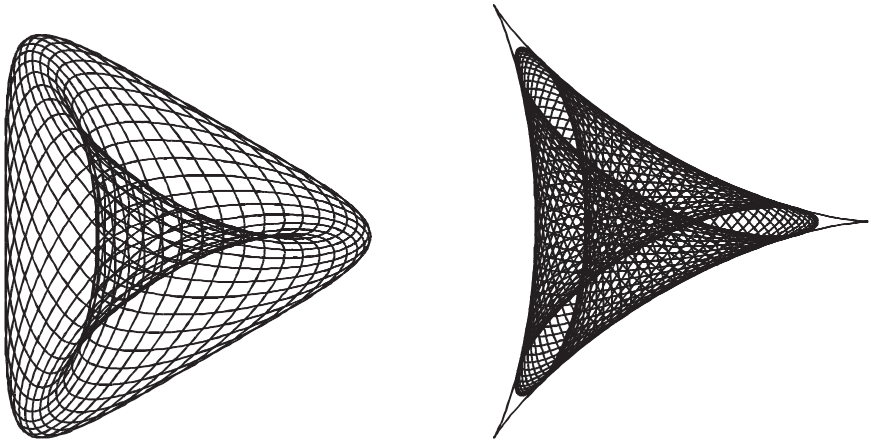


Figure 2 Collections of Simson's ellipses. The left-hand diagram is for $R = 2$ and the right-hand diagram is for $R = 0.7$. (The basic envelope is also drawn.)

Figure 2 indicates the plausibility of the envelope being in two parts. The starting point for finding the envelope of a family of curves given by $x = F(t, \phi)$ and $y = G(t, \phi)$ is to realise (or see http://mathworld.wolfram.com/Envelope.html) that the envelope must satisfy these two equations together with

$$\frac{\partial F}{\partial t}\frac{\partial G}{\partial \phi} = \frac{\partial G}{\partial t}\frac{\partial F}{\partial \phi}.$$

After a nonnegligible algebraic exercise, the envelope is indeed found to consist of two separate curves. One of them is exactly the curve given by (3), the basic envelope. This is most surprising. Figure 2 shows that the basic envelope is the inner of the two curves which are appearing for $R > 1$ and the outer curve when $R < 1$. The other curve that forms the envelope can be written as

$$x = x_0 + R\cos\phi + \tfrac{1}{2}\cos 2\phi, \qquad y = y_0 + R\sin\phi - \tfrac{1}{2}\sin 2\phi.$$

Again, the equations may be combined by using polar coordinates as in (2) to give

$$1 - \cos 3\psi = \frac{[4R^2 - (2r-1)^2](4R^2 - 2r - 1)^2}{64R^2r^3}. \tag{5}$$

The constraints $-1 \le \cos 3\psi \le 1$ imply that $|R - \frac{1}{2}| \le r \le R + \frac{1}{2}$. The cases $R > 1$ and $R < 1$ produce curves of a rather different form, and this is illustrated in figures 3 and 4. The case $R = \frac{1}{2}$ is worthy of separate mention because the envelope is then $r = \cos 3\psi$.

Now consider the family of envelopes—does this system itself have an envelope? So far as I am aware there is not a name for the envelope of a set of envelopes—how about a wrapper? The different envelopes for values of R exceeding unity (such as shown in figure 3) do not intersect and so do not have a wrapper. But, for R less than unity (see figure 4), it is clear that they all touch the basic envelope and so this forms their wrapper (as may be verified analytically).

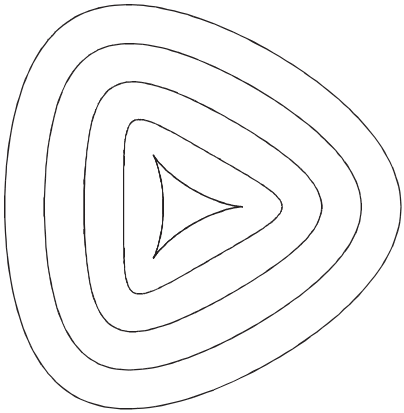


Figure 3 A collection of Simson's envelopes. The values of R are $1, 2, 3, 4, 5$.

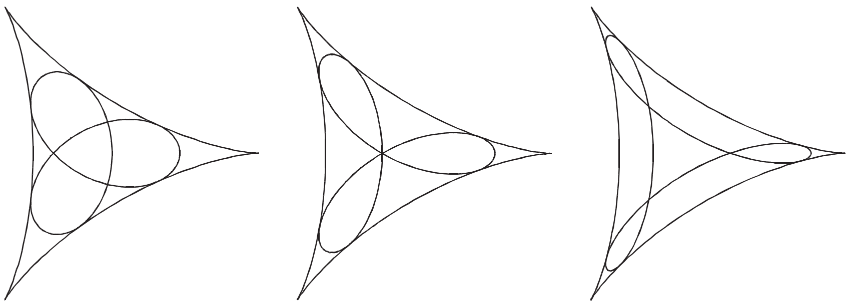


Figure 4 A collection of Simson's envelopes. The values of R are $0.3, 0.5, 0.7$. In each case the basic envelope is also shown.

5. Simson's axes

Instead of looking at the whole ellipse, let us now concentrate upon its axes. The equations of the major and minor axes are

$$(y-y_0)\cos\left(\frac{\phi}{2}\right)+(x-x_0)\sin\left(\frac{\phi}{2}\right)-\frac{1}{2}R\sin\left(\frac{3\phi}{2}\right)=0$$

$$\text{and}\quad (y-y_0)\sin\left(\frac{\phi}{2}\right)-(x-x_0)\cos\left(\frac{\phi}{2}\right)+\frac{1}{2}R\cos\left(\frac{3\phi}{2}\right)=0.$$

The envelopes of these axes are just scaled copies of the basic envelope (for the minor axis, it is rotated through $\pi/3$).

Denote the midpoint of PH by P'. Then P' is the centre of Simson's ellipse and so both axes pass through P'. Furthermore, NP' is parallel to OP and has length $\frac{1}{2}R$. As P moves around a circle of radius R centred on O, P' moves around a circle of radius $\frac{1}{2}R$ centred on N. If the point Q lies on OP, then the axes of the ellipse associated with Q are parallel to those associated with P. Hence, these axes can be constructed by ruler and compass methods for any point in the plane. However, the coordinate axes being used, for which (1) holds, cannot be constructed by such methods.

The endpoints of the major axis are

$$x = x_0 + \frac{1}{2}R\cos\phi \pm \frac{1}{2}(R+1)\cos\left(\frac{\phi}{2}\right), \quad y = y_0 + \frac{1}{2}R\sin\phi \mp \frac{1}{2}(R+1)\sin\left(\frac{\phi}{2}\right),$$

and the locus of these points using the polar coordinates of (2) is

$$1 - \cos 3\psi = \frac{(1+2r)(2R-2r+1)(1+2R-2Rr)^2}{16r^3R(R+1)^2}.$$

The extreme values for r are $\frac{1}{2}$ and $R+\frac{1}{2}$. For the minor axis, the corresponding results are

$$x = x_0 + \frac{1}{2}R\cos\phi \pm \frac{1}{2}(R-1)\sin\left(\frac{\phi}{2}\right), \qquad y = y_0 + \frac{1}{2}R\sin\phi \pm \frac{1}{2}(R-1)\cos\left(\frac{\phi}{2}\right),$$

and

$$1 - \cos 3\psi = \frac{(1+2r)(2R+2r-1)(1-2R+2Rr)^2}{16r^3R(R-1)^2},$$

the new extreme values of r being $\frac{1}{2}$ and $|R-\frac{1}{2}|$. Figure 5 shows a set of major axes. It will be noticed that a surprising number of them appear to be concurrent in threes. This is because if $\phi_1 + \phi_2 + \phi_3 = 0$ then the corresponding three major axes are concurrent. This result was noted in reference 1 as Property 9. A similar result holds for the minor axes (although this is not so clear) which are illustrated in figure 6. Note that these figures are not to their true scale; each has been made approximately the same size.

The case $R = \frac{1}{2}$ is worthy of special mention. The point P now lies on a circle with the same radius as the nine-point circle. The locus of the end points of the minor axis is then $2r = -\cos 3\psi$.

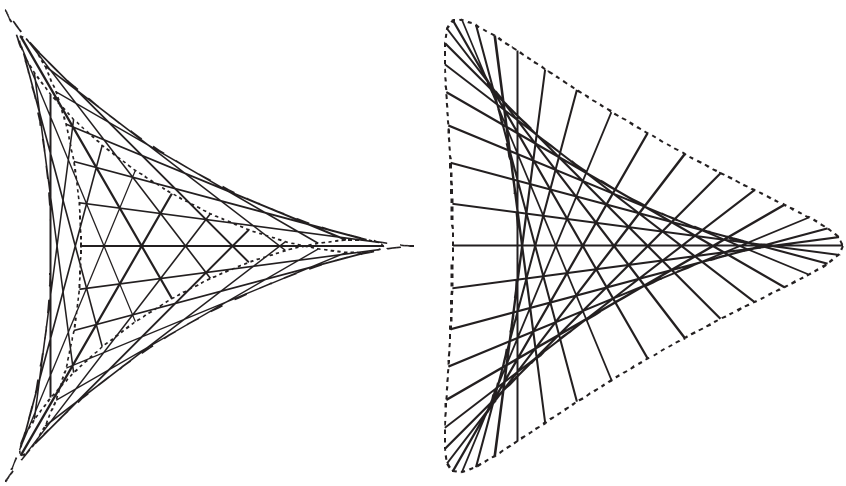


Figure 5 A collection of major axes. The left-hand diagram is for $R = 1.5$ and the other for $R = 0.5$. In both, the locus of the endpoints is shown as a dotted line. The basic envelope is shown as a dashed line in the left-hand diagram and forms (most of) the outer boundary. In the right-hand diagram it appears as the boundary to the cross-hatched region.

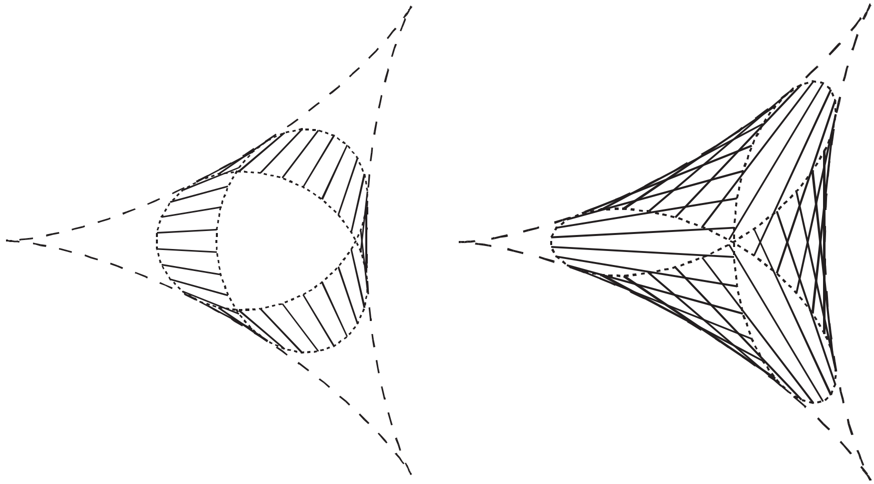


Figure 6 This is similar to figure 5 except that it is for minor axes.

For the classical situation ($R = 1$), it could be argued that *Simson's line* should refer to just the line segment whose length equals the diameter of the circumcircle and whose centre is the midpoint of PH. It is this line segment which has the curious property that its endpoints trace out a curve (the basic envelope) which is also the envelope of the segments. At the same time, the locus of the midpoint of these segments (which is now the same as the endpoints of the minor axis) is the nine-point circle.

6. Appendix. The derivation of the equation of Simson's ellipse

The slope of the line BC is

$$\begin{aligned}\frac{\sin\theta_2 - \sin\theta_3}{\cos\theta_2 - \cos\theta_3} &= \frac{2\cos(\theta_2/2 + \theta_3/2)\cdot\sin(\theta_2/2 - \theta_3/2)}{-2\sin(\theta_2/2 + \theta_3/2)\cdot\sin(\theta_2/2 - \theta_3/2)} \\ &= -\cot\left(\tfrac{1}{2}\theta_2 + \tfrac{1}{2}\theta_3\right) \\ &= \cot\left(\tfrac{1}{2}\theta_1\right).\end{aligned}$$

Thus, the equation of the line BC is $y - \sin\theta_2 = \cot(\theta_1/2)(x - \cos\theta_2)$ and of the line PD it is $y - R\sin\phi = -\tan(\theta_1/2)(x - R\cos\phi)$. The elimination of y yields

$$\begin{aligned}x &= -\tfrac{1}{2}\sin\theta_2\sin\theta_1 + \tfrac{1}{2}R\sin\phi\sin\theta_1 + \cos^2\left(\tfrac{1}{2}\theta_1\right)\cos\theta_2 + R\sin^2\left(\tfrac{1}{2}\theta_1\right)\cos\phi \\ &= -\tfrac{1}{2}\sin\theta_2\sin\theta_1 + \tfrac{1}{2}R\sin\phi\sin\theta_1 + \tfrac{1}{2}\cos\theta_2(1+\cos\theta_1) + \tfrac{1}{2}R\cos\phi(1-\cos\theta_1) \\ &= \tfrac{1}{2}[R\cos\phi - R\cos(\theta_1+\phi) - \cos\theta_1] + \tfrac{1}{2}[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)] \\ &= x_0 + \tfrac{1}{2}[R\cos\phi - R\cos(\theta_1+\phi) - \cos\theta_1].\end{aligned}$$

Now take t as a parameter and write

$$\begin{aligned}x &= x_0 + \tfrac{1}{2}[R\cos\phi - R\cos(t+\phi) - \cos t] \\ &= \left(x_0 + \tfrac{1}{2}R\cos\phi\right) - \tfrac{1}{2}(R\cos\phi + 1)\cdot\cos t + \tfrac{1}{2}R\sin\phi\cdot\sin t.\end{aligned}$$

A similar process gives

$$\begin{aligned} y &= y_0 + \tfrac{1}{2}[R\sin\phi + R\sin(t+\phi) - \sin t] \\ &= \big(y_0 + \tfrac{1}{2}R\sin\phi\big) + \tfrac{1}{2}R\sin\phi\cdot\cos t + \tfrac{1}{2}(R\cos\phi - 1)\cdot\sin t. \end{aligned}$$

So $t = \theta_1$ gives the point D, and likewise $t = \theta_2, \theta_3$ will give the points E, F, respectively.

Since the coordinates of P are $(R\cos\phi, R\sin\phi)$ and those of the orthocentre H are $(2x_0, 2y_0)$, those of the midpoint of PH are

$$\big(x_0 + \tfrac{1}{2}R\cos\phi,\ y_0 + \tfrac{1}{2}R\sin\phi\big).$$

If the origin is moved to this midpoint, then the equations for the locus have the form

$$x = p\cos t + q\sin t \quad\text{and}\quad y = r\cos t + s\sin t, \tag{6}$$

from which may be obtained $(ps - qr)\cos t = sx - qy$ and $(ps - qr)\sin t = -rx + py$. Hence,

$$(ps - qr)^2 = (sx - qy)^2 + (-rx + py)^2,$$

which is the equation of an ellipse (since x and y are bounded). Also, from (6) we have

$$\begin{aligned} x^2 + y^2 &= (p^2 + r^2)\cos^2 t + 2(pq + rs)\cos t\sin t + (q^2 + s^2)\sin^2 t \\ &= \tfrac{1}{2}(p^2 + q^2 + r^2 + s^2) + \tfrac{1}{2}(p^2 - q^2 + r^2 - s^2)\cos 2t + (pq + rs)\sin 2t. \end{aligned}$$

If a, b are the lengths of the semi-major and semi-minor axes, then the maximum and minimum values of $(x^2 + y^2)$ are a^2 and b^2, respectively. Thus,

$$a^2, b^2 = \tfrac{1}{2}(p^2 + q^2 + r^2 + s^2) \pm \tfrac{1}{2}[(p^2 - q^2 + r^2 - s^2)^2 + 4(pq + rs)^2]^{1/2}.$$

The substitution of $p = -\frac{1}{2}(R\cos\phi + 1)$, $q = r = \frac{1}{2}R\sin\phi$, $s = \frac{1}{2}(R\cos\phi - 1)$ into this formula yields the answers

$$a = \tfrac{1}{2}(R + 1) \quad\text{and}\quad b = \tfrac{1}{2}|R - 1|,$$

which are quoted at the beginning of section 4.

References

1 G. T. Vickers, Simson's envelope, *Math. Spectrum* **42** (2009/2010), pp. 5–13.

2 G. T. Vickers, Simson's triangle, *Math. Spectrum* **42** (2009/2012), pp. 42–43.

3 C. Jordan, Demonstrating the envelope of Simson's line, available at http://ms.appliedprobability.org//data/files/vickers/simsonenvelope.html.

Glenn Vickers *graduated from Sheffield University and then gained a PhD from Queen Mary College, London in astrophysics. He is now retired from the Applied Mathematics Department in Sheffield where he taught a wide range of topics including genetics, galactic structure, and evolutionary game theory.*

UK Personal Income: an Application of the Pareto Distribution

JOHN C. B. COOPER

The underlying theory of the Pareto distribution is explained and an empirical application using personal income data for the UK is presented to demonstrate estimation of the Pareto exponent.

Introduction

The Pareto distribution is named after Vilfredo Pareto, who was born in Paris in 1848 of Italian–French parents. He originally studied mathematics and physics prior to becoming a civil engineer, but later developed an intense interest in political economy. He was appointed Professor of Economics at the University of Lausanne in 1893 and his treatise on political economy (see references 1 and 2) confirms that he regarded economics as essentially a mathematical science.

One of Pareto's many contributions to economics is his observation that the distribution of income across a population follows a regular pattern with many individuals at the lower end of the income scale and *progressively* fewer at the upper end. This finding has proven remarkably resilient to empirical testing across different countries and time periods and has come to be known as Pareto's law of income distribution. Accordingly, the empirical example used in this article fits a Pareto distribution to UK personal income for the relatively recent fiscal years 1999/2000 and 2009/2010.

Theory

The probability density function (PDF) of the Pareto distribution may be written

$$f(x) = \alpha x_0^{\alpha} x^{-(\alpha+1)},$$

where $x > x_0,\ x_0 > 0$, and $\alpha > 0$. The distribution is continuous and positively skewed with the parameter x_0 representing some minimum value that the random variable may assume. The other parameter α, known as the *Pareto exponent*, characterises the decay of the distribution and must be estimated from a sample of data.

The cumulative distribution function (CDF) for a Pareto random variable in the range $x > x_0$ may be derived by integration in the usual way. Thus,

$$\begin{aligned} F(x_i) = \Pr(x_0 < x < x_i) &= \int_{x_0}^{x_i} \alpha x_0^{\alpha} x^{-(\alpha+1)}\, \mathrm{d}x \\ &= 1 - \alpha x_0^{\alpha} x_i^{-\alpha}, \end{aligned}$$

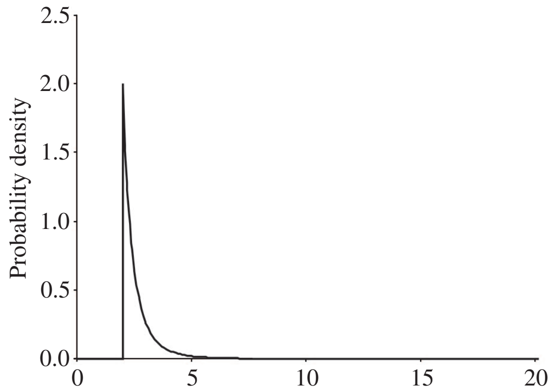


Figure 1 The Pareto distribution PDF.

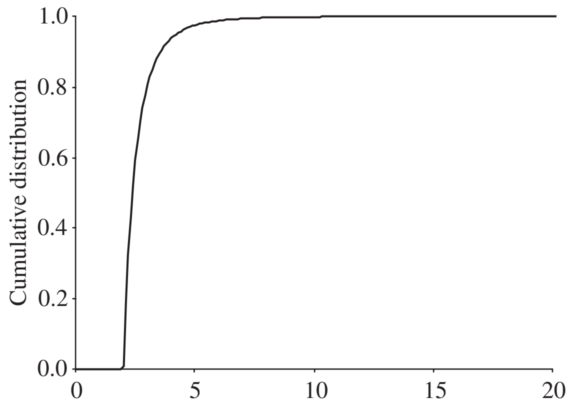


Figure 2 The Pareto distribution CDF.

so that

$$[1 - F(x_i)] = \Pr(x > x_i) = \alpha x_0^{\alpha} x_i^{-\alpha}. \tag{1}$$

Figures 1 and 2 show the PDF and CDF for a Pareto distribution with $x_0 = 2$ and $\alpha = 4$.

A useful property of (1) is that it may be readily linearised as

$$\ln[1 - F(x_i)] = \alpha \ln x_0 - \alpha \ln x_i. \tag{2}$$

Expressed in this form, (2) states that the logarithm of the proportion of units exceeding some value x_i is a negatively sloped, linear function of the logarithm of x_i. This negative gradient is of course the Pareto exponent which may be estimated using ordinary least squares (OLS).

Application

Pareto observed that, *above a certain minimum level of income*, the logarithm of the proportion of a population earning an income in excess of some amount was a negatively sloped linear

Table 1 UK personal income 1999/2000 (source: HM Revenue & Customs).

Income (£'000s), x_i	$\ln x_i$	Actual proportion with income greater than x_i, $[1 - F(x_i)]$	$\ln[1 - F(x_i)]$	Theoretical proportion with income greater than x_i, $x_0^{\alpha} x_i^{-\alpha}$
20	2.9957	1.0000	0.0000	1.0000
30	3.4012	0.4163	−0.8764	0.4467
50	3.9120	0.1287	−2.0503	0.1618
70	4.2485	0.0627	−2.7694	0.0829
100	4.6052	0.0303	−3.4966	0.0408
200	5.2983	0.0084	−4.7795	0.0103
500	6.2146	0.0017	−6.3771	0.0017

Table 2 UK personal income 2009/2010 (source: HM Revenue & Customs).

Income (£'000s), x_i	$\ln x_i$	Actual proportion with income greater than x_i, $[1 - F(x_i)]$	$\ln[1 - F(x_i)]$	Theoretical proportion with income greater than x_i, $x_0^{\alpha} x_i^{-\alpha}$
20	2.9957	1.0000	0.0000	1.0000
30	3.4012	0.5377	−0.6205	0.4685
50	3.9120	0.1645	−1.8048	0.1803
70	4.2485	0.0734	−2.6118	0.0961
100	4.6502	0.0312	−3.4673	0.0493
200	5.2983	0.0092	−4.6886	0.0135
500	6.2146	0.0032	−5.7446	0.0024

function of the logarithm of that amount. In other words, (2) fitted the distribution of income *at upper levels* but was not successful in explaining incomes at lower levels. Interestingly, Klein (see reference 3) confirmed this observation for UK incomes in the fiscal year 1953/1954.

Tables 1 and 2 show, for two different fiscal years, the distribution of UK personal income at several different levels x_i, together with the proportion of taxpayers earning in excess of these amounts, $[1 - F(x_i)]$. The figure of £20 000 was selected as the minimum value of x_i because it was the modal income for both years.

A regression of $\ln[1 - F(x_i)]$ on $\ln x_i$ produced the estimated equations

$$\ln[1 - F(x_i)] = 5.8030 - 1.9876 \ln x_i, \qquad R^2 = 0.996,$$

for 1999/2000, and

$$\ln[1 - F(x_i)] = 5.5008 - 1.8699 \ln x_i, \qquad R^2 = 0.985,$$

for 2009/2010. Clearly, the coefficient of determination R^2 indicates that the equation fits the data very well for both years.

Another way of establishing how well the Pareto distribution fits the income data is to compare the actual proportion of taxpayers earning in excess of x_i per annum (third column) with the theoretical proportion computed from (1) shown in the fifth column. For example, the theoretical proportion earning in excess of £100 000 per annum during 2009/2010 is calculated from

$$x_0^{\alpha} x_i^{-\alpha} = 20^{1.8699} \times 100^{-1.8699} = 0.0493.$$

This is very close to the actual value of 0.0312. A comparison of the third and fifth columns would suggest that the Pareto distribution provides a reasonable fit to the data for both fiscal years.

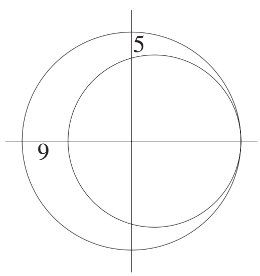
Finally, in this particular application to income distribution, the Pareto exponent α is actually a measure of the degree of income inequality. The lower the value of α, the higher the degree of income inequality. To see this, consider, for example, that when α equals 1.8797, the proportion of taxpayers earning in excess of £100 000 per annum is 4.93%. Had α equalled 2.8797, this proportion would have fallen to 0.99%. With this in mind, it is interesting to note that α has fallen from 1.9894 to 1.8797 between the two years of this small study, indicating that the contrast between the richest and poorest has intensified over time. (Interestingly, Klein (see reference 3, page 152) estimated α to be 1.981 for the fiscal year 1953/1954.)

References

1 V. Pareto, *Cours d'Economie Politique, Volume 1* (University of Lausanne, 1896).
2 V. Pareto, *Cours d'Economie Politique, Volume 2* (University of Lausanne, 1897).
3 L. R. Klein, *An Introduction To Econometrics* (Prentice Hall, New Jersey, 1962).

***John Cooper** is a senior lecturer in Financial Economics at Glasgow Caledonian University and holds visiting professorships in the USA, Peru, and Hungary. His research interests include the application of mathematical and statistical methods in financial decision-making.*

Touching circles



A small circle touches a larger circle internally as shown. Rectangular axes pass through the centre of the larger circle. The smaller circle cuts off 9 cm and 5 cm from the x-axis and y-axis, respectively. Find the diameters of the two circles.

Reference

1 R. Colson, *Perplexing Puzzles, Cryptic Challenges, Remarkable Riddles* (Parragon Book Services Ltd, Bath, 2012).

Midsomer Norton, Bath **Bob Bertuello**

Is the IPL 2012 Tournament a Fair Game to All Quarter-finalists? A Statistical Overview

JYOTI SHIWALKAR and M. N. DESHPANDE

The Indian Premier League, known as the IPL tournament, is a popular cricket match series among cricket lovers. The format for the seminal and final round of the IPL 2012 tournament was changed this time. This article aims to bring out the 'probabilistic' unfairness of the rules set for the semi-final and final rounds of the IPL 2012 tournament.

1. Introduction

In the recent 2012 Indian Premier League (IPL 2012), nine participating teams played in 72 league matches. Out of these nine teams, the four top teams were declared quarter-finalists on the basis of their performance in the league matches. The format for the final rounds was changed for this tournament and the changes are explained in the next section.

2. Pattern for the seminal and final rounds of IPL 2012

Suppose that A, B, C, and D are the top four teams respectively, and each qualified as quarter-finalist teams. Then the qualifier-1 match was played between A and B to decide the winner of qualifier-1. Next, the eliminator match was played between C and D to decide the loser and this team was knocked out. The qualifier-2 match was then played between the *loser* of qualifier-1 (A or B) and the *winner* of the eliminator (C or D). The final match was then played between the winner of qualifier-1 and the winner of qualifier-2. In all, four matches were played to decide the winner of IPL 2012.

3. Probabilistic unfairness of the game

Surprisingly, we found that the pattern used for deciding the winner in IPL 2012 is statistically not fair. This is because the chances of winning for teams A and B are three times greater than that of teams C and D. Furthermore, teams C and D share an equal chance of winning the tournament. This can be explained as follows.

Since the four quarter-finalist teams can be considered to be of equal standard, we assume that the probability of winning a match against any other quarter-finalist team is $\frac{1}{2}$. We evaluate the winning probability according to the above rule for each of the four quarter-finalist teams A, B, C, and D. In the next section P denotes the probability, and Q1 and Q2 refer to the qualifier-1 and qualifier-2 matches, respectively.

4. Calculation of winning probabilities

We have

$$
\begin{aligned}
\text{P[team A wins the tournament]} &= \{\text{P[A wins Q1]} \times \text{P[A wins the final after winningQ1]}\} \\
&\quad + \{\text{P[A loses Q1]} \times \text{P[A wins Q2 after losing Q1]} \\
&\quad \times \text{P[A wins the final after losing Q1 and winning Q2]}\} \\
&= \left\{\tfrac{1}{2} \times \tfrac{1}{2}\right\} + \left\{\tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2}\right\} \\
&= \tfrac{3}{8}.
\end{aligned}
$$

Similarly, P[team B wins the tournament] $= \frac{3}{8}$. However,

$$
\begin{aligned}
\text{P[team C wins the tournament]} &= \text{P[C wins the eliminator]} \\
&\quad \times \text{P[C wins Q2 after winning the eliminator]} \\
&\quad \times \text{P[C wins the final after winning eliminator and Q2]} \\
&= \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} \\
&= \tfrac{1}{8}.
\end{aligned}
$$

Similarly, P[team D wins the tournament] $= \frac{1}{8}$.

It clearly indicates that the rules of the IPL 2012 tournament give teams A and B a three times greater chance of winning than teams C and D, which is statistically unfair.

5. Concluding remarks

In almost all tournaments, two semi-finals are played between any of the two quarter-finalist teams, and the final match is then played between the winning teams of the two semi-finals. This most commonly used rule assigns an *equal chance* of winning to all the four quarter-finalist teams. Whereas in IPL 2012 there were four matches played in the semi-final and final rounds of the tournament, assigning *unequal chances* of winning to the quarter-finalist teams. Hence we can say that the IPL 2012 tournament was not a fair game to the quarter-finalists.

Jyoti Shiwalkar *is an Associate Professor in the Department of Statistics, Hislop College, Nagpur, India. Her research is in the field of probability and probability distributions.*

M. N. Deshpande *retired as Professor and Head of Department of the Department of Statistics, Institute of Science, Nagpur, India. He has been a doctoral adviser to numerous PhD candidates in the field of probability, probability distributions, and sampling techniques.*

When the Proof is Not Quite Enough

CHRISTINE SHANNON

This article presents a problem motivated by an experiment in psychology which is a nice application of both the conditional probability and infinite series. Furthermore, it provides an opportunity to use computer simulation first to build intuition about the correct solution to the problem and then to build confidence in the results once the theoretical calculations are complete. The problem is to select a sequence of pairs of color words (with replacement) from a set of four colors in such a way that there would be approximately equal numbers of matching and nonmatching pairs subject to the constraint that successive pairs must be distinct.

1. Introduction

Solutions to probability problems can easily hide subtle errors. The 'Monty Hall problem' which garnered much attention several years ago is a good case in point (see reference 1). Computer simulations can serve as an excellent vehicle both to guide our intuition in formulating a proof or solution and to convince us of the accuracy of our results at the end. A mathematical proof should stand on its own, but psychologically it is often comforting to see that what the theory predicts is, in fact, what happens in practice. When confronted with a question about which little is known, a quick simulation can often provide some intuition for either formulating a theorem for eventual proof or questioning an incorrect hypothesis. Similarly, when the solution of a problem requires a complicated sequence of steps, a simulation can bolster confidence in the calculations especially when approximations are required. This interplay between theoretical results and simulations can be very satisfying and instructive for students and professionals alike. This article describes one such case study.

2. An interdisciplinary application

Interesting mathematical questions often present themselves when least expected. There is a well known test in psychology called the Stroop task which is designed to assess how well the subject pays attention (see reference 2). In a particular version the color words 'red', 'yellow', 'green', and 'blue' are presented on a computer screen in one of the four colors and the subject is asked to respond with the name of the font color, which may or may not be the same as the color word being presented. For example, the word 'blue' might be presented in a red font and the subject should respond 'red'. The stimulus can be considered as a pair, ('color word', font color). If the word matches the color in which it is presented, then the pair is called *congruent* and if they do not, it is labeled as *incongruent.* The experiment consists of presenting a sequence of these stimuli to a subject and recording the average response time and accuracy for both congruent and incongruent pairs. The protocol calls for random choices of word and font color with approximately fifty percent being congruent pairs.

3. The basic algorithm

The algorithm for generating a random sequence of such pairs $\{(w_n, f_n)\}$ satisfying these constraints is fairly straightforward. Assume that `selectColor()` is a function that randomly returns a string from the set {'red', 'yellow', 'green', 'blue'} with a uniform distribution and that `random()` is a function which returns a real number uniformly distributed on the interval [0,1]. Then the following algorithm (written in a simple pseudocode) will meet the requirements:

```
selectPair()
   w ← selectColor()
   r ← random()
   if (r < 0.5)
      f ← w
   else
      f ← selectColor()
      while (f = w)
         f ← selectColor()
   return (w,f)
```

Half the time `f` and `w` will be equal and the rest of the time they will be different. A simulation readily supports the theory.

4. A slight change

It can be confusing to the subject if two identical pairs of color word and font are presented successively in a row, resulting in only one response when two are expected. Consequently, the goal is to generate a sequence $\{(w_n, f_n)\}$ satisfying all the original conditions but also the added requirement that, for all natural numbers n, at least one of the following is true: $w_{n+1} \neq w_n$, $f_{n+1} \neq f_n$. Except for the first pair where the original `selectPair()` will work, subsequent pairs will have to be chosen with a knowledge of the previous pair. Hence, it is necessary to define a second function which now has as argument the ordered pair `(oldw,oldf)` that was generated the last time:

```
selectPair2((oldw, oldf))
   (w,f) ← selectPair()
   while ((oldw, oldf) = (w,f))
      (w,f) ← selectPair()
   return (w,f)
```

This function calls the original function until a pair different from the previous pair is generated. This will satisfy the constraint that the pairs are random and that successive pairs are distinct, but, since the events are no longer independent, it may no longer be true that there are approximately equal numbers of congruent and incongruent pairs. Indeed, 50 runs of a simulation to generate 1 000 000 such pairs resulted in an average of 477 282 sets of congruent pairs per run. One can easily reject the hypothesis that the probability of a congruent pair is 0.5. This experiment is labeled as *experiment 1* and will be considered again at a later time. The obvious question is: with what probability should congruent pairs in the function `selectPair()` be generated so that after rejecting sequential duplicates in `selectPair2(oldw,oldf)` the subject will still be presented with approximately equal numbers of congruent and incongruent pairs?

For all practical purposes, one could experimentally adjust the probability until acceptable results are obtained, but it is a splendid opportunity to dust off some conditional probability and derive the theoretical result.

5. Mathematical analysis

The first requirement is to define the probability space for the experiment of generating an ordered pair using `selectPair()`. Both the color word and the font color are chosen from a uniform distribution. However, the protocol calling for equal numbers of congruent and incongruent pairs means that the ordered pairs are not equally likely. For example, table 1 indicates that the probability of the pair (‘red’, red) is $\frac{1}{8}$ and the probability of the pair (‘red’, green) is $\frac{1}{24}$. It is easy to calculate these numbers since the four congruent pairs must appear half the time and the 12 incongruent pairs must also appear half the time. Within each category the possible outcomes are equally likely.

For each n, let

$$X_n = \begin{cases} 1 & \text{if the pair } (w_n, f_n) \text{ is congruent,} \\ 0 & \text{if the pair } (w_n, f_n) \text{ is incongruent.} \end{cases}$$

Then indeed, $\mathrm{P}(X_1 = 1) = 0.5$. However, to calculate $\mathrm{P}(X_n = 1)$ for $n > 1$ it is necessary to employ the conditional probability. Indeed, for $n > 1$,

$$\begin{aligned} \mathrm{P}(X_n = 1) &= \mathrm{P}(X_{n-1} = 0)\,\mathrm{P}(X_n = 1 \mid X_{n-1} = 0) + \mathrm{P}(X_{n-1} = 1)\,\mathrm{P}(X_n = 1 \mid X_{n-1} = 1) \\ &= [1 - \mathrm{P}(X_{n-1} = 1)]\,\mathrm{P}(X_n = 1 \mid X_{n-1} = 0) \\ &\quad + \mathrm{P}(X_{n-1} = 1)\,\mathrm{P}(X_n = 1 \mid X_{n-1} = 1). \end{aligned}$$

To calculate the conditional probabilities, first assume that $X_{n-1} = 0$. This means that the previous pair was not congruent. Because the new pair cannot duplicate the previous pair, it may be necessary for `selectPair2((oldw,oldf))` to call the `selectPair()` function multiple times. Observe that if it is called k times, it is because the initial $k-1$ calls resulted in an exact match with `(oldw,oldf)` and the final call produced a congruent pair. Let p_k be the probability that, assuming the previous pair was incongruent, k calls are necessary and result in a congruent pair. Since the previous pair was incongruent and the desired pair is congruent,

$$p_k = \left(\frac{1}{24}\right)^{k-1} \frac{1}{2} \quad \text{for } k \geq 1\,.$$

Table 1

	Font color			
Color word	red	yellow	green	blue
‘red’	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
‘yellow’	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{24}$
‘green’	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{24}$
‘blue’	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{8}$

Thus, the event of getting a congruent pair given that the previous pair was incongruent can happen either because the first call results in a congruent pair *or* the first call exactly matches the previous pair and the second call is congruent *or* the first two calls exactly match the previous pair and the third is congruent, etc. Indeed, there is an infinite numberof disjoint ways in which this event can occur and, thus, for $n > 1$, $\mathrm{P}(X_n = 1 \mid X_{n-1} = 0)$ is given by an infinite geometric series

$$\mathrm{P}(X_n = 1 \mid X_{n-1} = 0) = \sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} \left(\frac{1}{24}\right)^{k-1} \frac{1}{2} = \frac{1}{2}\left(\frac{1}{1 - 1/24}\right) = \frac{12}{23}.$$

Now, assume that $X_{n-1} = 1$. This time the previous pair was congruent. As above, it may be necessary to call `selectPair()` k times, because the initial $k-1$ calls resulted in an exact match of the previous pair and the final call produced a congruent pair that is different from the previous pair. Let q_k be the probability that, assuming that the previous pair was congruent, k calls are necessary and result in a congruent pair. Since the previous pair was congruent and the desired pair is congruent but different, we have $q_k = \left(\frac{1}{8}\right)^{k-1} \frac{3}{8}$ for $k \geq 1$. As above,

$$\mathrm{P}(X_n = 1 \mid X_{n-1} = 1) = \sum_{k=1}^{\infty} q_k = \sum_{k=1}^{\infty} \left(\frac{1}{8}\right)^{k-1} \frac{3}{8} = \frac{3}{8}\left(\frac{1}{1 - 1/8}\right) = \frac{3}{7}.$$

The recurrence relation is now $\mathrm{P}(X_1 = 1) = 0.5$ and, for $n > 1$,

$$\mathrm{P}(X_n = 1) = (1 - \mathrm{P}(X_{n-1} = 1))\frac{12}{23} + \mathrm{P}(X_{n-1} = 1)\frac{3}{7} = \frac{12}{23} + \frac{-15}{161}\,\mathrm{P}(X_{n-1} = 1).$$

Back substitution results in

$$\begin{aligned}
\mathrm{P}(X_n = 1) &= \frac{12}{23} + \frac{-15}{161}\,\mathrm{P}(X_{n-1} = 1) \\
&= \frac{12}{23} - \frac{15}{161}\frac{12}{23} + \left(-\frac{15}{161}\right)^2 \frac{12}{23} + \cdots + \left(-\frac{15}{161}\right)^{n-2} \frac{12}{23} \\
&\quad + \left(-\frac{15}{161}\right)^{n-1} \mathrm{P}(X_1 = 1) \\
&= \frac{12}{23}\left(\frac{1 - (-15/161)^{n-1}}{1 - (-15/161)}\right) + \left(-\frac{15}{161}\right)^{n-1} \mathrm{P}(X_1 = 1). \qquad (1)
\end{aligned}$$

Thus, $\mathrm{P}(X_n = 1)$ is not constant for all values of n and, in fact, involves a partial sum of a geometric series in which the ratio is negative. Consequently, $\mathrm{P}(X_n = 1)$ will oscillate around the limit as n increases. Furthermore, observe that

$$\lim_{n\to\infty} \mathrm{P}(X_n = 1) = \frac{12}{23}\left(\frac{1}{1 - (-15/161)}\right) = \frac{21}{44} = 0.477\,27\overline{27}. \qquad (2)$$

This value is interesting because it helps to explain the result of experiment 1 where an average of 477 282 congruent pairs per run in 50 runs of 1 000 000 pairs was reported. Indeed, if $Y_n = \sum_{i=1}^{n} X_i$ then Y_n is a random variable which equals the number of congruent pairs in a series of n trials. It follows from probability theory that

$$\mathrm{E}[Y_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \mathrm{P}(X_i = 1).$$

The limit in (2) and the size of the ratio in the geometric series guarantee that, except for some very small values of n, $\mathrm{P}(X_n = 1)$ will be very close to $0.477\,272\overline{7}$, which makes the resultsmathematical calculations have helped to explain the results of the simulation.

6. A simulation to test the theory

It is now possible to conduct a simulation to see how closely the distribution of congruent pairs in the kth position agrees with the predicted values in (1). The experiment generated 1 000 000 sequences of length 10, using `selectPair()` for the first pair and `selectPair2((oldw,oldf))` for the subsequent pairs. The fraction of times that the kth pair was congruent was recorded for each k between 1 and 10. This observed frequency is the value of $\mathrm{P}(X_k = 1)$ in the second column of table 2. The third column gives the value of this probability as computed with (1). The final column demonstrates the oscillatory nature of the convergence that was indicated in the calculation of $\mathrm{P}(X_n = 1)$. Many interesting pieces of information can be gleaned from this table. Since

$$\mathrm{E}[Y_{10}] = \sum_{i=1}^{10} \mathrm{E}[X_i] = \sum_{i=1}^{10} \mathrm{P}(X_i = 1),$$

the sum of the entries in the second column indicates that the average number of congruent pairs in the $1\,000\,000$ samples of size 10 is $4.793\,446$. Similarly, the sum of the entries in the third column shows the corresponding theoretical expected value to be $4.793\,517\,562$. More importantly, the third column suggests that the convergence to the limit in (2) is rapid. Indeed, for $n > 10$, $\mathrm{P}(X_k = 1)$ will be $0.477\,272\,727$ when rounded to eight decimal places and

$$\begin{aligned}\sum_{i=1}^{1\,000\,000} \mathrm{P}(X_i = 1) &= \sum_{i=1}^{10} \mathrm{P}(X_i = 1) + \sum_{i=11}^{1\,000\,000} \mathrm{P}(X_i = 1) \\ &= 4.793\,517\,562 + 999\,990 \times 0.477\,272\,727 \\ &= 477\,272.7.\end{aligned}$$

Note that this calculation gives a little more insight into the results of the simulation in experiment 1.

Table 2

n	Estimated value of $\mathrm{P}(X_n = 1)$	Theoretical value of $\mathrm{P}(X_n = 1)$ from (1)	Difference
1	0.500 311 00	0.500 000 000 00	0.000 311 00
2	0.474 997 00	0.475 155 279 50	−0.000 158 28
3	0.477 562 00	0.477 470 005 02	0.000 091 99
4	0.477 013 00	0.477 254 347 36	−0.000 241 35
5	0.476 972 00	0.477 274 439 69	−0.000 302 44
6	0.476 652 00	0.477 272 567 73	−0.000 620 57
7	0.477 271 00	0.477 272 742 14	−0.000 001 74
8	0.477 172 00	0.477 272 725 89	−0.000 100 73
9	0.478 234 00	0.477 272 727 40	0.000 961 27
10	0.477 262 00	0.477 272 727 26	−0.000 010 73

7. Making it come out even

If we desire to have approximately equal numbers of congruent and incongruent pairs then, clearly, it is necessary to adjust the probability with which congruent pairs are selected in the function `selectPair()`. Let the constant ϕ be this probability. Thus, the new version is

```
selectPair()
   w ← selectColor()
   r ← random()
   if (r < φ)
      f ←w
   else
      f ← selectColor()
      while (f = w)
         f ← selectColor()
   return (w,f)
```

The second function is unchanged but the probability table becomes table 3. The table now indicates that the probability of the pair ('red', red) is $\phi/4$ and that the probability of the pair ('red', green) is $(1-\phi)/12$ since this will produce congruent pairs with probability equal to ϕ. It then follows that $\mathrm{P}(X_1 = 1) = \phi$ and for $n > 1$,

$$\begin{aligned}
\mathrm{P}(X_n = 1 \mid X_{n-1} = 0) &= \sum_{k=1}^{\infty} p_k \\
&= \sum_{k=1}^{\infty} \left(\frac{1-\phi}{12} \right)^{k-1} \phi \\
&= \phi \frac{1}{1-(1-\phi)/12} \\
&= \phi \frac{1}{(11+\phi)/12} \\
&= \frac{12\phi}{11+\phi}
\end{aligned}$$

and

$$\mathrm{P}(X_n = 1 \mid X_{n-1} = 1) = \sum_{k=1}^{\infty} q_k = \sum_{k=1}^{\infty} \left(\frac{\phi}{4} \right)^{k-1} \frac{3\phi}{4} = \frac{3\phi}{4} \left(\frac{1}{1-\phi/4} \right) = \frac{3\phi}{4-\phi}.$$

Table 3

	Font color			
Color word	red	yellow	green	blue
'red'	$\phi/4$	$(1-\phi)/12$	$(1-\phi)/12$	$(1-\phi)/12$
'yellow'	$(1-\phi)/12$	$\phi/4$	$(1-\phi)/12$	$(1-\phi)/12$
'green'	$(1-\phi)/12$	$(1-\phi)/12$	$\phi/4$	$(1-\phi)/12$
'blue'	$(1-\phi)/12$	$(1-\phi)/12$	$(1-\phi)/12$	$\phi/4$

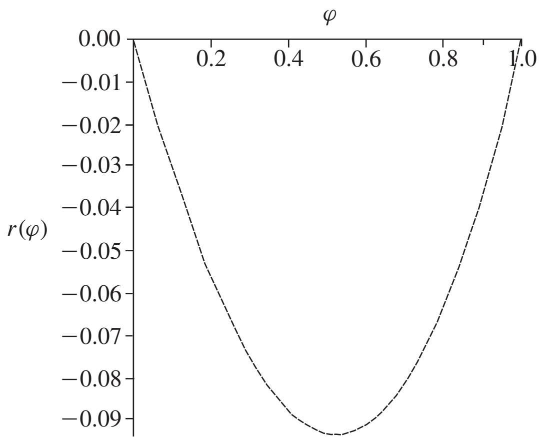


Figure 1

Finally, the recurrence relation becomes $\mathrm{P}(X_1 = 1) = \phi$ and, for $n > 1$,

$$\begin{aligned}
\mathrm{P}(X_n = 1) &= (1 - \mathrm{P}(X_{n-1} = 1))\frac{12\phi}{11+\phi} + \mathrm{P}(X_{n-1} = 1)\frac{3\phi}{4-\phi} \\
&= \frac{12\phi}{11+\phi} + \mathrm{P}(X_{n-1} = 1)\left(\frac{3\phi}{4-\phi} - \frac{12\phi}{11+\phi}\right) \\
&= \frac{12\phi}{11+\phi} + \mathrm{P}(X_{n-1} = 1)\frac{15\phi(\phi-1)}{(4-\phi)(11+\phi)}. \qquad (3)
\end{aligned}$$

Letting

$$r(\phi) = \frac{15\phi(\phi-1)}{(4-\phi)(11+\phi)},$$

it follows that $r(\phi) < 0$ and plotting $r(\phi)$ for ϕ in the interval $[0, 1]$ (see figure 1) reveals that the range of r is contained in the interval $[-0.1, 0]$. Thus, for some R in the interval $[-0.1, 0]$, the recurrence relation in (3) becomes $\mathrm{P}(X_1 = 1) = \phi$, and for $n > 1$,

$$\mathrm{P}(X_n = 1) = \frac{12\phi}{11+\phi} + \mathrm{P}(X_{n-1} = 1)R.$$

Using back substitution as before, it follows that

$$\begin{aligned}
\mathrm{P}(X_n = 1) &= \frac{12\phi}{11+\phi} + R\left(\frac{12\phi}{11+\phi} + \mathrm{P}(X_{n-2} = 1)R\right) \\
&= \frac{12\phi}{11+\phi} + R\left(\frac{12\phi}{11+\phi} + R\left(\frac{12\phi}{11+\phi} + \mathrm{P}(X_{n-3} = 1)R\right)\right) \\
&= \frac{12\phi}{11+\phi}(1 + R + R^2 + \cdots + R^{n-2}) + R^{n-1}\,\mathrm{P}(X_1 = 1) \\
&= \frac{12\phi}{11+\phi}\frac{1-R^{n-1}}{1-R} + R^{n-1}\phi.
\end{aligned}$$

Given the fact that R will be in the interval $[-0.1, 0]$, it is clear that, except for the first few values of n, $\mathrm{P}(X_n = 1)$ will be very close to

$$\lim_{n\to\infty}\left(\frac{12\phi}{11+\phi}\frac{1-R^{n-1}}{1-R} + R^{n-1}\phi\right) = \frac{12\phi}{11+\phi}\frac{1}{1-R}.$$

Substituting for R in terms of ϕ and setting this limit to 0.5 results in the equation

$$\frac{3\phi(4-\phi)}{11+2\phi-4\phi^2} = 0.5.$$

Solving for ϕ in the interval $[0, 1]$ gives $\phi = 0.525\,062\,81$. This means that if congruent pairs are selected with probability 0.525 062 81, then long sequences of pairs should contain approximately equal numbers of congruent and incongruent pairs.

8. Simulation results

This calculation involves an approximation. In order to simplify the calculation of ϕ, we replace $\mathrm{P}(X_n = 1)$ by the limit for all n. In this case, a simulation can investigate whether this approximation is too imprecise and if it is necessary to use the actual first few terms of the sum rather than the limit in each case. When the Stroop task is actually administered, it will probably present about 100 stimuli to each subject. As a very simple check of the appropriateness of the estimate, 200 sequences were generated with 100 pairs in each sequence and the number of congruent pairs was recorded. The results showed an average of 50.28 congruent pairs per 100 pair sequence with a standard deviation of 4.51. The researcher will have to decide if a tighter estimate is required.

Again, the individual distribution of the X_i can also be examined. We again ran 1 000 000 trials of ten pairs each, recorded the observed frequency of congruent pairs at each point in the sequence, and compared the results to the theoretical probabilities given in (3). Table 4 gives the results when we now select congruent pairs with $\phi = 0.525\,062\,81$.

As before, adding the values in column two gives the estimated value for the number of congruent pairs in a sequence of ten trials as $\mathrm{E}[Y_{10}] = \sum_{i=1}^{10} \mathrm{E}[X_i] = 5.022\,748$ and the sum of the third column gives the corresponding theoretical value as 5.022 921 9.

Table 4

n	Estimated value of $\mathrm{P}(X_n = 1)$	Theoretical value of $\mathrm{P}(X_n = 1)$ from (3)	Difference
1	0.524 704	0.525 062 81	−0.000 358 81
2	0.497 905	0.497 659 121	0.000 245 88
3	0.500 887	0.500 218 634	0.000 668 37
4	0.499 366	0.499 979 575	−0.000 613 58
5	0.499 643	0.500 001 903	−0.000 358 9
6	0.499 615	0.499 999 818	−0.000 384 82
7	0.500 899	0.500 000 013	0.000 898 99
8	0.499 928	0.499 999 994	−0.000 071 99
9	0.500 749	0.499 999 996	0.000 749
10	0.499 052	0.499 999 996	−0.000 948

9. Conclusion

Access to high-speed computers is almost ubiquitous and resorting to some simulations to reinforce theoretical results can be very reassuring. In the case described here, altering a simple experiment to guarantee that successive pairs of randomly generated words would not be identical led to some interesting probability questions involving the conditional probability and infinite series. Simulations were used first to raise a suspicion that congruent and incongruent pairs were not equally likely and then to support the theory and verify the suitableness of an estimate once mathematical calculations of the probabilities were completed. This kind of interplay between theoretical analysis and simulation can both inspire the direction of a proof or approximation and support its correctness. It can be that extra assurance when the proof 'is not quite enough!'

References

1 R. De Young, Stroop effect: a test of the capacity to direct attention, http://snre.umich.edu/eplab/demos/st0/stroopdesc.html (2011).

2 S. Lucas, J. Rosenhouse and A. Schepler, The Monty Hall problem, reconsidered. *Math. Magazine* **82** (2009), pp. 332–42.

Christine Shannon *has taught both mathematics and computer science for over 20 years at Centre College where she enjoys the interplay between the two fields. Seeing a nice application of mathematics when writing a programming application is a source of great delight.*

Letters to the Editor

Dear Editor,

The maximum distance between successive primes

The Russian mathematician Pafnuty Chebycheff (1821–1894) proved that a closed interval $[x, 2x]$, where x is a positive integer, contains at least one prime number.

Let p be a prime number, and suppose that there is no other prime in the interval $[p, 2p]$. Now consider the interval $[(p+1), 2(p+1)]$. This interval contains two numbers not included in the previous interval, namely, $2P + 1$ and $2p + 2$. Since $2p + 2$ is even, $2p + 1$ must be prime. Thus, if p is prime, the next prime is at most $2p + 1$.

Yours sincerely,

John Sherrill
(P.O. Box 5122
Lubbock TX 79408
USA)

Dear Editor,

A trajectory for maximum impact

I write in response to John Gilder's invitation at the end of his interesting article in Volume 44, Number 2, pp. 61–63.

Retaining the notation and numbering of the equations there, and with a target at elevation α and coordinates $(d, d\tan\alpha)$ with respect to O, the condition for the shell to hit the target after time t is $d = ut\cos\theta$ together with $d\tan\alpha = ut\sin\theta - \frac{1}{2}gt^2$. Eliminating t gives $d = 2u^2/g(\tan\theta - \tan\alpha)\cos^2\theta$ as the replacement for Equation (1) in Gilder's article. Using Equation (2), $mu^2 = b$, and with $I = mu\cos\theta$, we obtain

$$I^2 = \frac{b^2}{u^2}\cos^2\theta = \frac{2b^2}{gd}(\tan\theta - \tan\alpha)\cos^4\theta.$$

Differentiating to maximise I^2 (and, hence, I) gives the equation

$$\cos^2\theta - 4(\tan\theta - \tan\alpha)\cos^3\theta\sin\theta = 0,$$

leading to $4(\tan\theta - \tan\alpha)\cos\theta\sin\theta = 1$ for the maximising value of θ. This can be rewritten as

$$\cos 2\theta + \tan\alpha\sin 2\theta = \tfrac{1}{2}$$

or $\cos(2\theta - \alpha) = \frac{1}{2}\cos\alpha$, so that

$$\theta = \tfrac{1}{2}\alpha + \tfrac{1}{2}\cos^{-1}\left(\tfrac{1}{2}\cos\alpha\right).$$

This is the value of θ that maximises the shell's horizontal momentum. If we imagine the target as being on a uniform slope of elevation α, an analogous problem would be to maximise the component of momentum parallel to the slope at impact, but this looks considerably messier.

Yours sincerely,

Nick Lord
(Tonbridge School
Kent TN9 1JP
UK)

Dear Editor,

An interesting algebraic inequality

In 2009 Mihaly Bencze and Zhao Changjian (see reference 1) showed that

$$\frac{1 + x^n}{2} \ge \frac{1 + x + x^2 + \cdots + x^n}{n+1} \ge \left(\frac{x+1}{2}\right)^n \qquad \text{for any } x \ge 0. \tag{1}$$

The right-hand side of (1) can be proved by induction as follows. For $n = 1$, we have equality. Let us suppose inductively that

$$\frac{1 + x + x^2 + \cdots + x^k}{k+1} \ge \left(\frac{x+1}{2}\right)^k.$$

Applying the formula $x^r - 1 = (x-1)(x^{r-1} + \cdots + x + 1)$ for the natural numbers $k+1-i$, i with $i \le k$, we get

$$(x^i - 1)(x^{k+1-i} - 1) = (x-1)^2(x^{i-1} + x^{i-2} + \cdots + x + 1)(x^{k-i} + x^{k-i-1} + \cdots + x + 1) \ge 0.$$

Therefore,

$$x^{k+1} + 1 \ge x^i + x^{k+1-i}, \tag{2}$$

so that

$$\begin{aligned}
k(x^{k+1} + 1) &\ge \sum_{i=1}^{k} (x^i + x^{k+1-i}),\\
k(x^{k+1} + 1) &\ge 2(x + x^2 + \cdots + x^k),\\
(2k+2)(x^{k+1} + 1) - 2(x + x^2 + \cdots + x^k) &\ge (k+2)(1 + x^{k+1}),\\
(2k+2)(1 + x + x^2 + \cdots + x^{k+1}) &\ge (k+2)[1 + 2(x + x^2 + \cdots + x^k) + x^{k+1}],\\
\frac{1 + x + x^2 + \cdots + x^{k+1}}{k+2} &\ge \frac{(x+1)(1 + x + x^2 + \cdots + x^k)}{2(k+1)}\\
&= \frac{x+1}{2}\,\frac{1 + x + x^2 + \cdots + x^k}{k+1}\\
&\ge \left(\frac{x+1}{2}\right)^{k+1}
\end{aligned}$$

by the inductive hypothesis. This proves the inductive step.

The left-hand side of (1) can be proved as follows. Substituting $k+1 = n$ into (2) we get $x^n + 1 \ge x^i + x^{n-i}$. Therefore,

$$(n+1)(x^n + 1) \ge \sum_{i=0}^{n} (x^i + x^{n-i}) = 2(1 + x + \cdots + x^n).$$

Hence,

$$\frac{1 + x^n}{2} \ge \frac{1 + x + \cdots + x^n}{n+1}.$$

Reference

1 M. Bencze and Z. Changjian, A refinement of Jensen's inequality, *Octogon Math. Magazine* **17** (2009), pp. 209–214.

Yours sincerely,

Spiros P. Andriopoulos
(Third High School of Amaliada
Eleia
Greece)

Dear Editor,

$$S_a + S_b = S_c$$

In Volume 44, Number 2, p. 50, Tom Moore considered the equation $T_a + T_b = T_c$, where T_a is the ath triangular number. I wondered about a similar equation for the numbers

$$S_n = 1^2 + 2^2 + \cdots + n^2.$$

Now

$$S_{n-1} + n^2 = S_n.$$

Try putting $S_k = n^2$. Then $k(k+1)(2k+1) = 6n^2$. One solution of this equation is $k = 24$, $n = 70$, so that

$$S_{69} + S_{24} = S_{70}.$$

Are there any other solutions?

I have found one solution of the equation $S_x = T_y$, namely, $S_5 = T_{10}$. Are there others apart from the obvious $S_1 = T_1$?

The following identities hold:

$$S_{6k} = kT_{12k+1}, \qquad S_{6k-1} = kT_{12k-2}, \qquad \frac{S_k}{T_k} = \frac{2k+1}{3}.$$

Yours sincerely,
Abbas Rouholamini
(Sirjan
Iran)

Dear Editor,

Conjectures about triangular numbers

In Volume 45, Number 2, p. 86, I proposed the conjecture that the only solutions of the equation

$$T_x + T_y = z^2,$$

where x, y, z are positive integers and $T_x = \frac{1}{2}x(x+1)$ denotes the xth triangular number, occur when $x = y + 1$ or $y = x + 1$. (We note that $T_x + T_{x-1} = x^2$.) I now realize that this conjecture is false.

We first note that the equation can be rewritten as

$$(2x+1)^2 + (2y+1)^2 = (2z+1)^2 + (2z-1)^2.$$

If we compare this with the identity

$$(am+bn)^2 + (an-bm)^2 = (an+bm)^2 + (am-bn)^2,$$

we put

$$2x+1 = am+bn, \qquad 2y+1 = an-bm,$$
$$2z+1 = an+bm, \qquad 2z-1 = am-bn,$$

so the equation is satisfied if we can choose positive integers a, b, m, n such that

$$am+bn, \qquad an-bm, \qquad an+bm$$

are positive odd integers and $(an+bm) - (am-bn) = 2$. For example, let $a = 31$, $b = 3$, $m = 6$, and $n = 5$. This gives

$$T_{100} + T_{68} = 86^2.$$

Yours sincerely,
Abbas Rouholamini
(Sirjan
Iran)

Problems and Solutions

Students are invited to submit solutions to some or all of the problems below. The most attractive solutions received by 1st March will be published in a subsequent issue and are eligible for annual prizes. When writing to the Editorial Office, please state your full name and also the postal address of your school, college, or university.

Problems

46.1 A uniform nondegenerate triangular lamina ABC has its centre of mass at G. The lamina is now removed and A, B, and C are joined by uniform rods to create a triangle. The centre of mass is now at H. If the point G coincides with the point H, show that ABC is an equilateral triangle.

(Submitted by Jonny Griffiths, Paston College, Norfolk)

46.2 For a given natural number n, how many irrational numbers of the form $\sqrt{m}$ are there smaller than n, where m is a natural number?

(Submitted by Bablu Chandra Dey, Kolkata, India)

46.3 For positive real numbers a, b, c with $a \leq b \leq c$, prove that

(i) $\dfrac{a}{b} + \dfrac{b}{c} + \dfrac{c}{a} \geq \dfrac{a}{c} + \dfrac{c}{b} + \dfrac{b}{a}$,

(ii) $b^2(c^2 - a^2)(c - a)^2 \geq a^2(c^2 - b^2)(c - b)^2 + c^2(b^2 - a^2)(b - a)^2$.

(Submitted by Spiros Andriopoulos, Third High School of Amaliada, Eleia, Greece)

46.4 If $P(x)$, $Q(x)$, and $R(x)$ are polynomials such that

$$xP(x^3) + Q(x^3) = (x^2 + x + 1)R(x),$$

prove that $x - 1$ is a factor of $P(x)$.

(Submitted by Anand Kumar, Patna, India)

Solutions to Problems in Volume 45 Number 2

45.5 Prove that

$$\int_0^{\pi/4} \mathrm{e}^{\tan x} \cos x \, \mathrm{d}x > 1.$$

Solution by Henry Riccardo, Tappan, NY

The curve $y = \mathrm{e}^x$ is convex and so lies above its tangent line $y = 1 + x$ at the origin, so that

$$\begin{aligned}\int_0^{\pi/4} \mathrm{e}^{\tan x} \cos x \, \mathrm{d}x &> \int_0^{\pi/4} (1 + \tan x) \cos x \, \mathrm{d}x \\ &= \int_0^{\pi/4} (\cos x + \sin x) \, \mathrm{d}x \\ &= [\sin x - \cos x]_0^{\pi/4} \\ &= 1.\end{aligned}$$

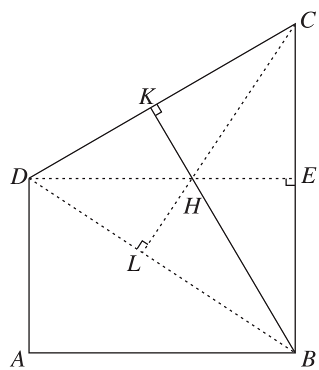
Also solved by Bor-Yann Chen, University of California, Irvine; Mr/Ms Hasler, France; Subramanyam Durbha, Community College of Philadelphia, USA; Kuntal Chaterjee; and Juan-Pablo Alonso and Angel Plaza, Spain. Spiros Andriopoulos, who posed the problem, uses the inequality

$$\mathrm{e}^{\tan x} > 1 + \tan x + \frac{(\tan x)^2}{2} + \frac{(\tan x)^3}{3}$$

to prove that the integral is greater than 1.1 . Henry Riccardo points out that a computer algebra system gives the value as approximately 1.1119.

45.6 The trapezium $ABCD$ has $\angle ABC = \angle BAD = 90°$, $AD < BC$, and K is the orthogonal projection of B onto CD. Prove that $DK = DA$ if and only if $CD = CB$.

Solution by Michel Bataille, who proposed the problem.



The intersection H of the lines DE and BK as shown is the orthocentre of $\triangle BCD$ and the right-angled triangles HKD and HEB are similar. Hence,

$$\begin{aligned} DK = DA \quad &\Longleftrightarrow \quad DK = EB \\ &\Longleftrightarrow \quad \triangle\text{s } HKD \text{ and } HEB \text{ are congruent} \\ &\Longleftrightarrow \quad HD = HB \\ &\Longleftrightarrow \quad DL = LB \\ &\Longleftrightarrow \quad CD = CB, \end{aligned}$$

where L is the foot of the perpendicular from C to BD.

Also solved by Subramanyam Durbha, Community College of Philadelphia, USA.

45.7 The points A, B on the ellipse with equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

are such that AB is perpendicular to the x-axis, and the point P with coordinates $(m, 0)$ is such that A, B, and P are not collinear. The line PB meets the ellipse again at E. Show that AE passes through the point $(a^2/m, 0)$.

Solution by Zhang Yun, who proposed the problem

The equation of PB can be written as

$$y = k(x - m),$$

and this meets the ellipse when

$$b^2x^2 + a^2k^2(x - m)^2 = a^2b^2,$$

i.e. $(b^2 + a^2k^2)x^2 - 2a^2k^2mx + a^2(k^2m^2 - b^2) = 0$. Denote the coordinates of B by (x_1, y_1), so that A has coordinates $(x_1, -y_1)$, and E by (x_2, y_2). Then

$$x_1 + x_2 = \frac{2a^2k^2m}{b^2 + a^2k^2}, \qquad x_1x_2 = \frac{a^2(k^2m^2 - b^2)}{b^2 + a^2k^2}.$$

The line AE has equation

$$\frac{y - y_2}{-y_1 - y_2} = \frac{x - x_2}{x_1 - x_2},$$

and this crosses the x-axis when $y = 0$, i.e. when

$$\begin{aligned}
x &= x_2 + \frac{x_1 - x_2}{y_1 + y_2} y_2 \\
&= x_2 + \frac{x_1 - x_2}{k(x_1 - m) + k(x_2 - m)} k(x_2 - m) \\
&= x_2 + \frac{(x_1 - x_2)(x_2 - m)}{(x_1 + x_2) - 2m} \\
&= \frac{x_1x_2 + x_2^2 - 2mx_2 + x_1x_2 - m(x_1 - x_2) - x_2^2}{(x_1 + x_2) - 2m} \\
&= \frac{2x_1x_2 - m(x_1 + x_2)}{(x_1 + x_2) - 2m} \\
&= \frac{2a^2(k^2m^2 - b^2) - 2a^2k^2m^2}{2a^2k^2m^2 - 2m(b^2 + a^2k^2)} \\
&= \frac{a^2}{m}.
\end{aligned}$$

45.8 Find a simple expression for the infinite product

$$P = (1 + x)(1 + x^2)(1 + x^4) \cdots (1 + x^{2^{n-1}}) \cdots,$$

where $|x| < 1$. If P_n is the nth partial product, how are $P - P_{n+1}$ and $P - P_n$ related?

Solution by Juan-Gabriel Alonso and Angel Plaza, Garoé Atlantic School and Universidad de Las Palmas de Gran Canaria, Spain.

$$P_n = \prod_{k=0}^{n-1}(1 - x^{2^k}).$$

Therefore, $P = \lim_{n\to\infty} P_n$. Note that

$$\begin{aligned}
(1-x)P_n &= (1-x)(1+x)(1+x^2)(1+x^4)\cdots(1+x^{2^{n-1}}) \\
&= (1-x^2)(1+x^2)(1+x^4)\cdots(1+x^{2^{n-1}}) \\
&= (1-x^4)(1+x^4)\cdots(1+x^{2^{n-1}}) \\
&= (1-x^{2^{n-1}})(1+x^{2^{n-1}}) \\
&= 1 - x^{2^n}.
\end{aligned}$$

So

$$P_n = \frac{1-x^{2^n}}{1-x},$$

and then

$$P = \lim_{n\to\infty} P_n = \lim_{n\to\infty} \frac{1-x^{2^n}}{1-x} = \frac{1}{1-x}.$$

On the other hand,

$$P - P_{n+1} = \frac{1}{1-x} - \frac{1-x^{2^{n+1}}}{1-x} = \frac{x^{2^{n+1}}}{1-x},$$

while

$$P - P_n = \frac{x^{2^n}}{1-x}$$

and, therefore, $P - P_{n+1} = x^{2^n}(P - P_n)$.

Also solved by Spiros Andriopoulos, Third High School of Amaliada, Eleia, Greece.

Infinite Roots

$$2 = \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}},$$
$$3 = \sqrt{6 + \sqrt{6 + \sqrt{6 + \cdots}}}.$$

Generally,

$$n = \sqrt{m + \sqrt{m + \sqrt{m + \cdots}}},$$

where $m = n^2 - n$.

Kolkata, India **Bablu Chandra Dey**

Mathematical Spectrum

Volume 46 2013/2014 Number 1

ISSN 0025-5653

http://ms.appliedprobability.org

Published by the Applied Probability Trust

Printed by Berforts Information Press Ltd, Stevenage, Hertfordshire, UK