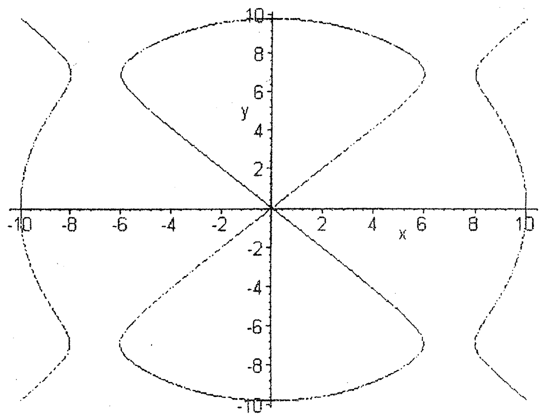
Function

A School Mathematics Journal

Volume 25 Part 5 October 2001



Department of Mathematics & Statistics – Monash University

Reg. by Aust. Post Publ. No. PP338685/0015

Function is a refereed mathematics journal produced by the Department of Mathematics & Statistics at Monash University. The journal was founded in 1977 by Prof G B Preston. *Function* is addressed principally to students in the upper years of secondary schools, and more generally to anyone who is interested in mathematics.

Function deals with mathematics in all its aspects: pure mathematics, statistics, mathematics in computing, applications of mathematics to the natural and social sciences, history of mathematics, mathematical games, careers in mathematics, and mathematics in society. The items that appear in each issue of *Function* include articles on a broad range of mathematical topics, news items on recent mathematical advances, book reviews, problems, letters, anecdotes and cartoons.

* * * * *

Articles, correspondence, problems (with or without solutions) and other material for publication are invited. Address them to:

The Editors, *Function*
Department of Mathematics & Statistics
PO BOX 28M
Monash University VIC 3800, AUSTRALIA
Fax: +61 3 9905 4403
e-mail: michael.deakin@sci.monash.edu.au

Function is published five times a year, appearing in February, April, June, August, and October. Price for five issues (including postage): $27.50* ; single issues $7.50. Payments should be sent to: The Business Manager, Function, Department of Mathematics & Statistics, PO Box 28M, Monash University VIC 3800, AUSTRALIA; cheques and money orders should be made payable to Monash University.

For more information about *Function* see the journal home page at http://www.maths.monash.edu.au/~cristina/function.html

---

* $14 for *bona fide* secondary or tertiary students.

The Front Cover

One of the classics of School Mathematics is the book *Mathematical Models* by Cundy and Rollett. These authors were both school teachers and their book first appeared in 1951. It deals with a wide variety of subjects, but perhaps of especial interest is the list of "miscellaneous curves" on pp 65-69.

Our Front Cover shows one of these. It is given by the equation

$$y^2\left(y^2-96\right)=x^2\left(x^2-100\right)$$

and because of its appearance, it is given the name "electric motor". A close look at the scale on the graph shows that only a part of the curve is shown. It has been cropped at $x = \pm 10$ and at $y = \pm 10$. This focuses on the central section, which resembles the armature of an electric motor fitting between the poles of the magnet that provides the field in which the armature rotates.

For values of x and y beyond these, the curve is less interesting. Readers should have no difficulty in showing that the curves $y = \pm x$ are asymptotes. (See the cover article to the previous issue.) So very soon beyond the closeup displayed on the cover, the curve "settles down" into a simpler pattern.

Back when Cundy and Rollett wrote, there were no simple means of producing curves like that shown. They had either to be laboriously plotted or else sketched by seeking their salient features and so trying to display the overall shape and properties of the curve, but without the onerous task of plotting.

Such skills are no longer cultivated, in part because they gave been rendered redundant by the availability of good quality graphics packages. The Front Cover was produced with the assistance of one such: MAPLE. A few lines of instructions are all that is required. Here they are.

```
>with(plots):
>implicitplot((y^2)*(y^2-96)=(x^2)*(x^2-100),x=-10..10,
y=-10..10,numpoints=10000,colour=black);
```

The computer does the rest!

Our cover for *Volume* **19**, *Part* **1** showed a stylised drawing of a drawbridge. It was based on one published by an eighteenth century military engineer, Bernard Forest de Bélidor. The illustration showed how the task of raising and lowering the drawbridge was made easier by the provision of a counterweight that ran up and down a carefully designed track. In an ideal case, the track turns out to be shaped like an arc of a curve known as a *cardioid*, although in practice this is merely an approximation. In the article accompanying the cover diagram, it was stated that the connection with the cardioid was discovered only recently.

This now seems not to be correct. There was an earlier analysis that gives precisely this result. There are also several other aspects of that earlier article that can now be updated. In the article that follows, Klaus Treitz gives the background and reviews the theory. A different version of this article first appeared in the German magazine *Mathematik in der Schule*, *Vol* **37** (1999), pp 30-33. What follows is an edited translation.

(Apfelmännchen, Little Apple Man, is the German name for a prominent feature of the Mandelbrot Set.)

APFELMÄNNCHEN AT KÖNIGSTEIN

Klaus Treitz, Basler Str, Rheinfelden, Germany

Fort Königstein lies in the Elbsandstein Mountains in the eastern part of Germany and is a well-known tourist destination. The website

http://www.festung-koenigstein.de/en

offers a first look and a virtual walk around the whole site. Apart from many impressive views, there is something very special for visitors interested in Mathematics. After crossing the first drawbridge and passing through the Medusa gate, you will discover two elegantly curved stone ramps behind the open doors. They begin just below the wheels supporting the ropes that were used to pull up the bridge. What is the purpose of these ramps?

It can easily be guessed. It was on these ramps that the counter-weights moved. The tracks were curved in such a way as to move the bridge up and down with as little effort as possible. How did the tracks have to be shaped for this purpose? The largest force is needed when you begin pulling the bridge up. The counterweights therefore must pull straight down in this situation. Later on, when the bridge is in an almost vertical position, the force required is almost zero. The counterweights need not to pull any longer. In other words, the tracks should be level at that point.

Ideally, the bridge should be in equilibrium throughout the lift; that is to say that it is in balance in all positions, with no tendency either to close or to open. Which curve ensures equilibrium in any position of the bridge? There are several different approaches to this problem.

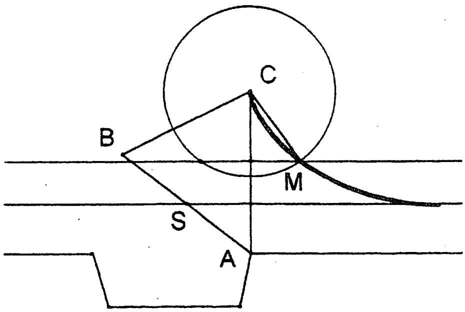


Figure 1

Figure 1 shows the bridge in a general position. AB represents the movable span, hinged at A. AC represents the tower supporting the rope BCM, whose right-hand point M gives the position of the counterweight, running on the curved track shown at the right. The rope is supposed to pass over a pulley situated at C. We assume that $AB = AC$. If the bridge were fully lowered with the weight then at the top of the tower, in that case M would coincide with C and the length of the rope BC would be $AB\sqrt{2}$. This length is to be constant as the bridge moves; call it a. Let m_B be the mass of the bridge and m_C the mass of the counterweight.

In the ideal case where there is no friction, the requirement of effortless mobility means that the energy of the bridge-counterweight system is to be constant. Suppose the centre of gravity is situated at a point S on AB. If the end of the rope between the pulley C and the counterweight M has the length r, then the length of the segment between the end B of the bridge and the pulley is $a - r$.

Using this background, we may draw the system of bridge, rope and counterweight with geometry software like Euklid. Thereafter it's a fun to open and close the bridge on the screen with the help of the dragging tool whilst the point M plots the required curve. The dynamic software Euklid is available as shareware from the website

http://www.dynageo/eng/index.html

The figures in this paper have been generated by this tool. Euklid itself provides descriptions of its constructions.

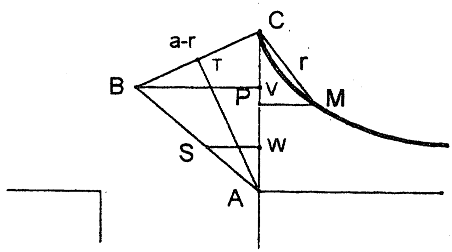


Figure 2

Suppose that in Figure 2 the distance of point B from the support A is k times the distance SA. Then from the geometry of the case where the bridge is fully down we have $a = AB\sqrt{2}$. Furthermore, by analysing the situation when the bridge is fully down it can be shown from basic mechanics that we must choose $m_C = \frac{m_B\sqrt{2}}{k}$.

Now get back to the general case of Figure 2. Because the triangles CVB and CTA are similar we have $\frac{CV}{CB} = \frac{CT}{CA}$ and therefore

$$CV = \frac{CB \times CT}{CA} \quad \text{i.e.} \quad CV = \frac{(a-r) \times \frac{1}{2}(a-r)}{\left(a/\sqrt{2}\right)} = \frac{(a-r)^2}{a\sqrt{2}}$$

Thus the point B will by lifted by the distance

$$AV = AC - CV = \frac{a}{\sqrt{2}} - \frac{(a-r)^2}{a\sqrt{2}} = \frac{1}{\sqrt{2}}\left(2r - \frac{r^2}{a}\right).$$

However the centre of gravity S is only lifted by a distance

$$AW = \frac{1}{k} AV.$$

On the other side the counterweight must move down a distance

$$CU = \frac{m_B}{m_C} AW = \frac{k}{\sqrt{2}} AW = \frac{k}{\sqrt{2}} \frac{1}{k} AV = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}}\left(2r - \frac{r^2}{a}\right) = \frac{1}{2}\left(2r - \frac{r^2}{a}\right)$$

We have thus found the position of M. If M is distant r from the pulley C, it must be $\frac{1}{2}\left(2r - \frac{r^2}{a}\right)$ below C. If we put $\angle PCM = \phi$, we then have

$$r = 2a(1 - \cos\phi). \qquad (*)$$

Note that this result is independent of the total mass and of the position of the centre of gravity, S, on AB.

Equation (*) is the equation, in polar co-ordinates, of a curve known as a cardioid (from the Greek, meaning “heart-shaped curve”). We may produce this curve in the following way (Figure 3): let one circle with radius a roll around another circle of the same size and follow a particular point of the first circle during the motion. The rolling condition ensures that the other two marked angles in the figure are also equal to ϕ.

From the isosceles trapezium $CMNK$ we get the polar equation (*) of the cardioid at a glance. Today this curve has become familiar as an especially prominent part of the Mandelbrot set. In German it is called

"Apfelmännchen" (little apple man). For the tracks supporting the counterweights only the small portion CE of the curve is needed.

This theory was developed by an early exponent of the calculus, the Marquis de l'Hôpital and his teacher, Johann I Bernoulli, one of a very famous family of mathematicians. Figure 4 is a copy of Bernoulli's drawing. l'Hôpital had found the curve in question for himself, but he did not recognise what it was. He asked Bernoulli to translate his solution into Latin and have it published in the learned journal *Acta Eruditorum.* This Bernoulli did in 1695, and he followed the analysis up with a geometric characterisation of the curve in question. The resulting papers are now reprinted in Johann I Bernoulli's *Opera Omnia* (Collected Works), Volume 1, pp 129-138. The correspondence between the two is reprinted in Volume 1 of Bernoulli's *Briefwechsel* (Correspondence), starting on p 244.

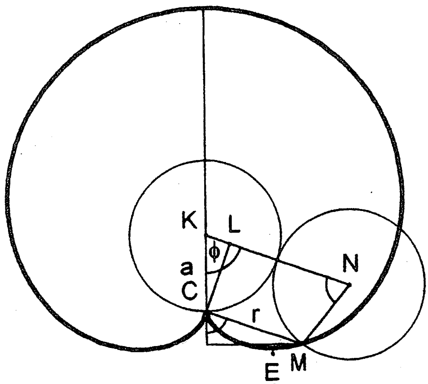


Figure 3

In the first of the two papers, Bernoulli presented the equation for the curve; in the second, he gave the geometric property that relates it to the rolling circle. He adopted some very clever simplifications, whose details are omitted here, but the result can easily be given. In Figure 4, put as he did $CP = x$, $PM = y$. (Note that this is somewhat different from our modern convention, where x represents a horizontal and y a vertical distance.) His analysis then resulted in the equation

$$a\sqrt{x^2 + y^2} = ax + \frac{1}{2}\left(x^2 + y^2\right),$$

which is the same as Equation (*).

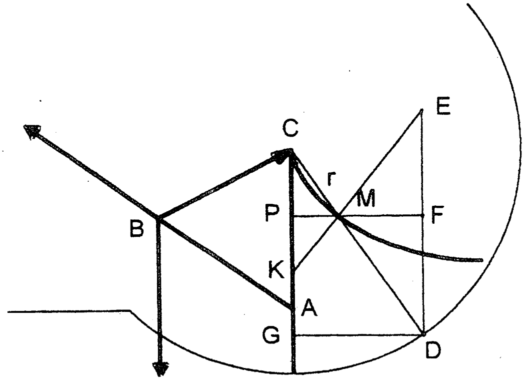


Figure 4

Nowadays we can use computer software to plot curves such as this, but back then things were different. So Johann Bernoulli was justifiably proud to be the first to recognise the nature of the curve in question: "I discovered, that this curve – and this is really noteworthy – belongs to the genus of cycloids .... described by a wheel rolling around another equal wheel". [The simplest cycloid for which one wheel rolls on a line was the subject of *Function's* cover story for February 1990. *Eds.*] Later on this particular curve was named the cardioid.

The oldest known bridge to be built along these lines was erected in the ninth century in Corsica. The huge ramps in the fortress of Bonifacio are marvellous but they look more like cosine curves than cardioids.

[In the 17th century fortress building flourished and different types of hinged drawbridges were developed. Other methods of setting up the required balance were also employed. One bridge built along these alternative lines can be seen at the fortress of Belfort near Basel, and another at the Port de Brisach, built by Sebastien de Vauban (1633-1707) a contemporary of de l'Hôpital.]

In the year 1684 Mathematics made an enormous step forward when Leibniz published his *Calculus Differentialis*. Now numerous curves could be investigated and problems of architecture and mechanics were attacked with the new tool. The question of the appropriate curves for the counterweights of a hinged drawbridge was raised by Joseph Sauveur, a friend of de l'Hôpital in the year 1694. He confessed to the Marquis that he had encountered great difficulties with the problem. l'Hôpital found the equation of the curve sought. But without the knowledge of trigonometric functions and polar coordinates it was difficult to recognise that the curve was a cardioid.

Bernoulli called the problem the "search for the *curva aequilibrationis*". The 1695/96 editions of the *Acta* contain no fewer than five articles on the problem, by l'Hôpital, Johann and Jakob Bernoulli and Leibniz. Leibniz expressed his admiration for these "most beautiful results". These matters are even mentioned in literature, namely in Laurence Sterne's satirical book *Tristram Shandy*. The name *cardioid* was coined by Johann Castillon (1704-1791) in 1741.

The Medusa gate of the Königstein was erected between 1730 and 1760. The curves of its bridge are quarter circles with a radius of 3.2 m. Unfortunately there exist no historical documents, but the plan worked out by a modern architect shows the great skill and knowledge of the builders and craftsmen in the 18th century

l'Hôpital and Johann Bernoulli thought along theoretical lines. They didn't pay attention to such complicating effects as friction. But when operating a real bridge we have to deal with a heavy construction which must be moved against large forces of friction. Circles instead of cardioids help to overcome this difficulty.

At the endpoints of a quarter circle the equilibrium condition is fulfilled, but halfway down the counterweights have a surplus of about 30% compared with the amount necessary. This helps to overcome frictional forces when lowering the bridge.

Such bridges were soon forgotten in Europe. But they experienced a renaissance in Australia and America. In particular, in New South Wales movable bridges were needed for roads crossing rivers which also were used for navigation and a number were built. They consisted of large steel or wooden frameworks and the curved tracks were composed of circular arcs with different radii. These curves were beautiful approximations of cardioids. Some of them still exist today.

[For more details on these Australian bridges, see the earlier Function *article. Eds.]*

The Mathematics of Elections

In the immediate aftermath of the controversial US presidential election that saw George W Bush narrowly defeat Al Gore comes a book on the Mathematics of elections. The author is Donald Saari, a US mathematician whose recent research has been devoted to just this topic. [The topic itself was briefly touched on in earlier issues of *Function* in the course of articles on C L Dodgson (Lewis Carroll). See *Function, Vol* **7**, *Part* **3** or *Vol* **18**, *Part* **1**. This is one of Dodgson's claims to fame as a mathematician, but in truth his pioneering efforts did not go very deeply into the matter.]

Saari's new book is called *Chaotic Elections*, and one of its points is that something very like chaos can emerge from various methods of deciding the outcome of elections. It also aims to draw attention to the difficult and challenging problems that the social sciences can pose for the mathematician. It goes deeply into its subject matter and makes for lively reading at the same time.

While much of it is understandably devoted to the bizarre complexities of the US system, there is a lot of a general nature, applying widely. Essentially, no system is perfect, but this does not gainsay the fact that some are better than others. When it comes to making recommendations of his own, Saari throws his weight behind a method first proposed by E J Nanson.

And here we have an Australian connection. Nanson was the second professor of Mathematics at the University of Melbourne, active in the last quarter of the nineteenth century. In 1882, he addressed the question of elections in a paper read to the Royal Society of Victoria, who later published by his analysis as a small pamphlet. The University of Melbourne holds a copy in their Mathematics library, and it makes interesting reading.

Suppose that several candidates all run for a single post. This is the case for seats in the lower house in Canberra today (but it does not apply to the Senate, which is more complicated). To simplify the discussion, suppose that three candidates are standing: **L** from the leftist party, **R** from the rightist party, and **C** from the centrist party.

What happens currently in Australia is that voters fill in ballot papers and put the numbers 1, 2, 3 opposite the names of these candidates, so indicating their order of preference. Thus one voter might well put a 1 opposite **L**, a 2 opposite **C** and a 3 opposite **R**. This would be a typical vote for a person of left-wing persuasion, who would probably prefer a centrist candidate to a right-wing one.

In the counting process, first of all, all the primary votes (1's) are counted. If one of the three has more than half of all the votes cast, then that candidate is clearly the winner, and there is no need to proceed any further. But often this is not the case. When no clear winner emerges from this first count, then the candidate with the *least* number of primary votes is eliminated. That candidate's second preferences are then allocated to the remaining two candidates and added to their existing scores. Thus, unless the totals happen to be exactly equal, then one of the two candidates will score more than the other and so will be elected.

Nanson attributed this method of deciding the outcome to W R Ware, a Mathematics professor from Harvard, but he did not like it and argued against it. The following example will show the source of his dissatisfaction. Suppose that **L** received 47% of the primary vote, and **R** 48%. The remaining 5% went to **C**. As no candidate has so far received 50+%, we must proceed to the next stage. **C** is declared defeated and the second preferences redistributed. Suppose that of the 5%, 80% (i e 4% of the total vote) went to **L** and the rest to **R**. Then **L** now has (47 + 4)% of the total and, as this is 51%, is duly declared elected.

But Nanson raised an objection that we can demonstrate in terms of this example. We may suppose that of the 47% voting for **L**, all or almost all would prefer **C** to **R**. For simplicity suppose it is all of them. Then the number preferring **C** to **R** is this 47% plus the 5% who directly voted for **C**; this gives us a total of 52% who prefer **C** to **R**. But similarly we find that a total of 53% prefer **C** to **L**! The compromise candidate, who was eliminated, would be most acceptable to the majority of voters!

Indeed we may think that something very like this happens in Australian elections, with the Australian Democrat candidate filling the place of **C**.

Nanson thus favoured a different system, and Saari agrees with him. This system would count every primary vote in a 3-candidate race as worth 2 points and every second preference as worth 1 point. Thus each voter allocates a total of 3 points. We can continue to talk in terms of percentages if we think of each primary vote as $\frac{2}{3}$%, and each second preference as $\frac{1}{3}$%.

[More generally, if there were n candidates standing, a primary vote would be worth $n - 1$ points, a second preference $n - 2$, etc all the way down to last place, which would be worth 0 points.]

In our example, **L** received 47% of the primary votes and a further 4% of second preferences. This then gives **L** a total percentage of

$$47 \times \frac{2}{3} + 4 \times \frac{1}{3} = 32.67\%$$

of the overall *points*. In the same way, **R** receives

$$48 \times \frac{2}{3} + 1 \times \frac{1}{3} = 32.33\% .$$

C's share is

$$5 \times \frac{2}{3} + 95 \times \frac{1}{3} = 33.33\% .$$

On this count, then, **R** is eliminated. This leaves **L** and **C**, and essentially we conduct a virtual runoff between them. **R**'s primary votes no longer count, and so (as in the system we in fact use) **R**'s second preferences count as primary votes. We now use a system as before, but with 2 candidates instead of 3. In this simple case, this does not affect the outcome: **C** wins over **L**, 53 to 47. (Had we started with a larger slate of candidates, things could get rather more interesting.)

The thing to notice about the Nanson method (or perhaps we should now call it the Nanson-Saari method) is that it does not really pick the most-liked candidate. In our example, **R** is the most popular of the three. What it does instead is to pick the *least-disliked* of the candidates. Whether this is seen as "fair" depends largely on your point of view; a lot of people would probably say not.

But after all that is hardly a mathematical question. Mathematics can only decide things when we first agree on what it is we want to achieve!

HISTORY OF MATHEMATICS

Benford's Law

Michael A B Deakin, Monash University

One of the strangest observations in the recent history of Mathematics was first effectively published by Frank Benford in 1938, although he himself said at the time that others had already noticed the same thing. (Several of the various websites devoted to the topic draw attention to an 1881 account by the American scientist Simon Newcomb. It was however Benford who brought the observation into general view.) Back in those days, before the arrival of the electronic calculator, it was customary to use tables of logarithms as computational aids. (For more detail on this, see *Function, Vol* **22**, *Part* **2**, p 57.) What Benford noticed was that, after a lot of use, tables of logarithms became much more dog-eared and smudged near the beginning than towards the end. This story is corroborated by Warren Weaver, author of *Lady Luck: The Theory of Probability.*

Now had the tables been dreary novels that failed to hold their readers' interest, then this is just what we would expect to see. Readers would get so far and then give up, some sooner, some later, but only rarely getting to the back of the book. But logarithmic tables?

What Benford deduced was that smaller numbers occurred more frequently in calculations than did larger ones. The tables he examined began with the number 1 000 000 and continued to 9 999 999, and of course, each of the initial digits 1, 2, ... , 9 takes up exactly one ninth of the total number of pages in the table-book. But clearly this simple equality was not reflected in the natural occurrence of the numbers in practice.

Indeed Benford investigated the matter and tabulated data from 20 sets of tables giving such data as the specific heats and molecular weights of thousands of chemical compounds, the surface areas(!) of 355 rivers, the street addresses of 342 persons listed in *American Men of Science*, and other such. All in all, he classified over 20 000 such numbers according to the first significant digit in the table where the number appeared. Then he calculated the frequency of each type, with the result given in Table 1.

The pattern strongly reminded Benford of another pattern, familiar to almost all those back then, who made regular use of logarithmic tables. This is given in Table 2.

INITIAL DIGIT	FREQUENCY
1	0.306
2	0.185
3	0.124
4	0.094
5	0.080
6	0.064
7	0.051
8	0.049
9	0.047

Table 1: Observed Frequencies as tabulated by Benford

INITIAL DIGIT	FREQUENCY	LOG FORMULA
1	0.306	0.301
2	0.185	0.176
3	0.124	0.125
4	0.094	0.097
5	0.080	0.079
6	0.064	0.067
7	0.051	0.058
8	0.049	0.051
9	0.047	0.046

Table 2: Comparison between Observed and Theoretical Frequencies

The third column is calculated by taking logarithms (to base ten) of each of the numbers (p) in the first column and then using the formula

$$f(p) = \log(p+1) - \log p \qquad (*)$$

(remembering that log 1 = 0).

Investigating further, Benford discovered that this formula worked better for some lists of numbers than it did for others. Thus it did well for the street addresses and for the areas of rivers, but it was not followed by the specific heats (which tend to lie close to 1) nor by the square roots of integers, which are generated by a known formula. Benford excluded such cases from further consideration and concentrated on those other numbers for which no theoretical basis could be given for their calculation. He called these numbers "anomalous numbers".

He was thus led to announce his "Law of Anomalous Numbers", according to which such numbers were distributed according to his formula because that was how the world was made; it was a Law of Nature like other logarithmic laws in Physics, Chemistry and Biology.

However, this was not the common view. Within a very few years, there were various attempts to explain *why* Benford's Law holds. Most such attempts fall into two main categories. One line of explanation starts from the presupposition that there is *some law or other* that gives the frequencies of the various initial digits, but then goes on to consider the effect of change of scale on the outcome. Thus Benford, in 1930s USA, would have used the old Anglo-Saxon system of weights and measures, whereas we in Australia today use SI units.

So, for example, if we tabulated the surface areas of rivers, one list might give the result in square miles, another in square kilometres, yet another in acres, etc. The numbers themselves would be different, but would the distribution of their initial digits remain the same?

This was the question addressed by the statistician Roger Pinkham in 1961. Pinkham proved that *if* there was an underlying law giving the frequencies of the initial digits, and if that law remained the same when the system of units involved was altered, then the underlying law *had* to be Benford's Law, i.e. Equation (*).

The other line of attack was more subtle, and its general outlines were indicated by Weaver in the book cited above. However, even before this, some correspondence in the British journal *Nature* had sought explanations along similar lines.

Consider the case of the street addresses. Now Pinkham's argument won't work for these because they are independent of our system of weights and measures. But street addresses are drawn from a finite list of integers. We in Australia have street numbers that often involve only two digits; less commonly, but not uncommonly, they run into the hundreds, and sometimes, just sometimes, they top the thousand mark. In America, it is quite common for street numbers to run to four or even five digits, but six is uncommon. Many other such lists could be described in similar terms, although the details would be different.

What Weaver pointed out was that if we put a "ceiling" on the size of the numbers we have available, then the frequency of the various initial digits can be calculated as a function of that ceiling (assuming that the various numbers up to and including the ceiling each occur with equal probability). Thus if 1 was the only number available, then 1 would *have* to be the initial digit. No other would be possible; and so we would have

$$f(1)=1,\ f(2)=f(3)=...=f(9)=0.$$

If we set the ceiling at 2, then the result would be

$$f(1)=f(2)=0.5, f(3)=f(4)=...=f(9)=0,$$

and so on until we reached a ceiling of 9, when we would have

$$f(1)=f(2)=f(3)=f(4)=...=f(9)=0.1111.... .$$

After this, the frequency of the initial digit 1 rises steadily until we reach a ceiling of 19, when we have

$$f(1)=0.5789..., f(2)=f(3)=f(4)=...=f(9)=0.0526.... .$$

And so we proceed. For each ceiling N, there is a calculable probability $f_N(p)$ that the initial digit will be p.

Weaver's idea was that if we let the value of the ceiling N tend to infinity, then there would be a limiting distribution and that this would be given by Benford's Law (*). He even gave a nice graphical argument to this effect. However, it was not precise enough to survive as a formal proof. What was needed was an account of a method of taking the mean of each frequency as N gets larger, in such a way that a sensible answer is produced.

Eventually this technical problem was overcome, by Mrs B J Flehinger of IBM. Her work was published in 1966, and it showed great ingenuity, although it was necessarily rather artificial. Her method of attack was to consider the average value of (first of all) $f_N(2)$, which she called A_N, as N increased. This value varies between 1 (when $N = 1$) and 0.111...., when $N = 9$, and it fluctuates, but staying between these values thereafter.

If we simply consider the means of all these values, we get other numbers B_N, whose values however also fluctuate, but a little less wildly. What she did next was to take the mean values of the B_N, to get a new set of averages C_N, which also fluctuated, but a little less wildly yet, and so she continued, and eventually showed that if the means were taken *infinitely often*, the answer would be log 2. A similar argument applied to each other initial digit.

This argument improved on the earlier analyses by the correspondents to *Nature*. Fortunately, however, this whole line of argument has since been greatly simplified. In 1969, another mathematician, R L Duncan, found a way to compute the averages much more simply, so that what took Mrs Flehinger so much time and effort could be done much more straightforwardly. Essentially what Duncan did was to put Weaver's somewhat imprecise argument on a firm footing, although he may not actually have known of that earlier work, as he made no reference to it.

At about this time, yet another attack was made on the problem. Another mathematician, Ralph Raimi, noted that "most of the time", geometric sequences, if carried far enough, gave results in accordance with Benford's Law. This version of matters has appeared once before in *Function*, and so here I will simply refer readers to Problem 3.3.5, whose solution appeared in *Volume* **5**, *Part* **3**, pp 16-18.

Raimi wrote a popular article on Benford's Law, and in this he was concerned (among other things) to show the superiority of the "Flehinger-type" explanation over the "Pinkham-type". One of his most telling arguments was that the "Pinkham-type" analysis does not apply to the case of the street addresses, whereas the "Flehinger-type" does. "Addresses .... are artifacts of man, not nature, and have nothing to do even with the British or metric systems." And elsewhere, "one would have to be a numerologist to make sense out of the operation of multiplying every entry in *Who's Who* by some positive constant".

I tend to agree with the general point that he makes, but nonetheless I took these remarks as a challenge. It seemed to me that although we might not use either system of weights and measures in producing integer numbers like street numbers, but all the same we can write such numbers in different bases. What if some other culture used a base other than ten?

So I was able to show that *if there is a law describing the distribution of first digits, and if that law is to depend on the base in which the numbers are written in a natural way*, then that law must be Equation (*), but modified so that we now write the logarithms in the new base b, rather than in base ten. Equation (*) uses base ten for the logarithms precisely because *we* use base ten in our counting system.

So, for example, if we used base 2, then the first digit would *necessarily* be 1, i.e. $\log_2 2$. If the base were 3, the first digit would be 1 with probability $\log_3 2 = 0.631$, etc. (Not dissimilar points were earlier advanced by the authors of the early correspondence in *Nature*, and this was a matter also taken up by R W Hamming in work described below.)

More recently, there has been renewed interest in Benford's Law. I will cite only a single example. In 1970, a very great mathematician, R W Hamming, who had introduced Pinkham to the possibility of research in the area, himself wrote on the topic. Hamming was interested in the efficiency with which numerical computations could be performed by computers. He went on to give several examples of the way in which a knowledge of Benford's Law could be put to good use in efficient computer programming. The details are technical and I will omit them here, but Hamming concluded that Benford's Law had practical application and was "not merely an amusing curiosity".

References

I have relied most particularly on Warren Weaver's book *Lady Luck: The Theory of Probability*, published by Heinemann as *Volume* **24** in their Science Study Series (see especially pp 270-277), and on Ralph Raimi's popular article "The Peculiar Distribution of First Digits" in

Scientific American (December 1969). Either of these sources should be quite easy for readers of *Function* to follow. More difficult are the technical articles listed below.

Frank Benford's original article is available in the *Proceedings of the American Philosophical Society, Volume* **78** (1938), pp 551-572. The early correspondence in *Nature* is the work of S A Goudsmit and W H Furry (*Volume* **154**, (1944) pp 800-801) and also of W H Furry and H Hurwitz (*Volume* **155**, (1945) pp52-53).

Roger Pinkham's result was published in *The Annals of Mathematical Statistics, Volume* **32** (1961), pp 1223-1230. Mrs Flehinger produced her ingenious argument in the journal *American Mathematical Monthly, Volume* **73** (1966), pp 1056-1061. Duncan's improvement on it is less easy to find; it appeared in *The Fibonacci Quarterly, Volume* **7** (1969), pp 474-475. My own work was the subject of a note in the *Australian Mathematical Society Gazette, Volume* **20** (1993), pp 162-163. Hamming published his account in the *Bell System Technical Journal, Volume* **49** (1970), pp 1609-1625.

ooooooooooooooooooooooooooooo

LETTER TO THE EDITOR

More on Pell's Equation

I would like to comment on Julius Guest's article "An Elementary Solution to Pell's Equation" in *Function* (June 2001).

Positive integer solutions to the second degree indeterminate equation

$$x^2 - Ny^2 = 1,$$

where N is not a perfect square, can be obtained by using the theory of continued fractions.

Briefly, when $\sqrt{N}$ is converted into an infinitely recurring continued fraction, the penultimate convergent in each of the recurring periods supplies a solution to the equation. Therefore there must be an infinite set of solutions. This topic is admirably dealt with in *Higher Algebra* by Hall & Knight. (I have a well-worn and much valued copy of the 1940 edition!)

Now to some comments on the particular equation chosen by Julius Guest,

$$x^2 - 3y^2 = 1.$$

By inspection, $x = 2$, $y = 1$ are the simplest solutions. We obtain all the others in the published list and as many more as we like by using the formula

$$\frac{x_{n+1}}{y_{n+1}} = \frac{1}{2}\left(\frac{x_n}{y_n} + N\frac{y_n}{x_n}\right)$$

In this case, $N = 3$ and therefore

$$\frac{x_2}{y_2} = \frac{1}{2}\left(\frac{2}{1} + 3\times\frac{1}{2}\right) = \frac{7}{4}$$

$$\frac{x_3}{y_3} = \frac{1}{2}\left(\frac{7}{4} + 3\times\frac{4}{7}\right) = \frac{97}{56}$$

and so on. Further, the general solution to

$$x^2 - 3y^2 = 1$$

is given by the two equations

$$x_n = \frac{1}{2}\left[\left(2+\sqrt{3}\right)^n + \left(2-\sqrt{3}\right)^n\right]$$

$$y_n = \frac{1}{2\sqrt{3}}\left[\left(2+\sqrt{3}\right)^n + \left(2-\sqrt{3}\right)^n\right]$$

for $n = 1, 2, 3, \dots$.

Assuming that $x = h$, $y = k$ are the "simplest" solutions for

$$x^2 - Ny^2 = 1$$

(found by inspection or otherwise), what are the equations representing the general solution?

David Shaw
Wellington St
Newtown, Vic

ooooooooooooooooooooooooooooooo

COMPUTERS AND COMPUTING

Winding Curves

Cristina Varsavsky

This column deals with a family of curves that appear very often in nature: the spirals. Spirals are usually formed during the growth process of organisms; they are present in some shells such as the nautilus shell, in spider webs, and in the heads of daisies. Spirals are also the building blocks of the living world, in the nucleic acid or DNA. But the most amazing and spectacular spirals are visible only through a telescope—these are the nebulas and galaxies of the universe.

A spiral is a curve that winds around a point and gradually recedes from that point. Here we will deal only with planar spirals, of which there are several types. Spirals are best described with polar coordinates. The relationship between the cartesian coordinates x and y of a point P in the plane and its corresponding polar coordinates r and θ is given by

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta, \end{aligned} \tag{1}$$

where r is the radial distance of the point P to the origin, and θ is the angle that the segment OP makes with the positive x–axis, as shown in Figure 1.

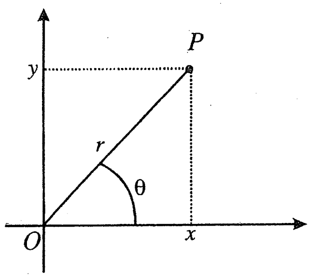


Figure 1

The simplest spiral is the *spiral of Archimedes*, defined as

$$r = a\,\theta\,. \tag{2}$$

Let us create a spreadsheet to find a few points on the curve. We put angle values, spaced by say 0.2 radians, in column **A**, then the formula (2) for r with $a = 100$ in column **B**, and the corresponding formulae (1) in columns **C** and **D** respectively. Finally we copy down to several rows the 3 formulae in **B**, **C** and **D**. Figure 2 shows the first few rows of the spreadsheet, and Figure 3 (overleaf) the graph of the points corresponding to the x and y values calculated in columns **C** and **D** joined together with a smooth line.

theta	r	x	y
0	0	0	0
0.2	20	19.60133	3.973387
0.4	40	36.84244	15.57673
0.6	60	49.52014	33.87855
0.8	80	55.73654	57.38849

Figure 2

You can play with different values of a, positive and negative, to see the effect this parameter has on the spiral curve. You will see that this type of spiral always starts at the origin and that the coils are always equidistant from each other. The spiral could be interpreted as the result of a uniform circular motion combined with a linear motion. Imagine a bug on a revolving music record crawling at a constant speed in a straight line from the centre of the record and towards the edge. Seen from above the record, the bug would describe a spiral of Archimedes.

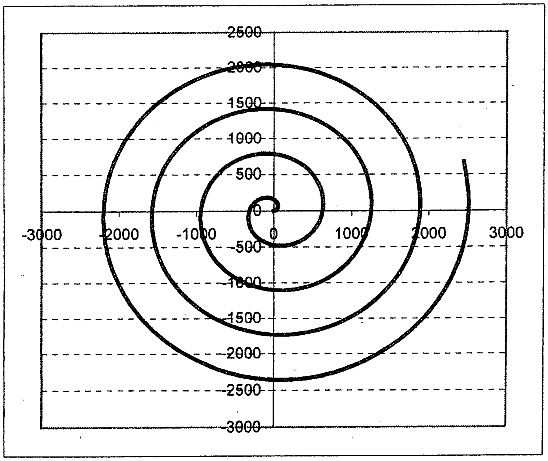


Figure 3

Formula (2) tells us that for every point on a spiral of Archimedes, r is proportional to θ. Therefore a spiral of Archimedes could be used to divide an arbitrary angle into any number of equal parts. For example, to trisect the angle POA in Figure 4, we draw circles centred at the origin with radii $\frac{r}{3}$ and $2\frac{r}{3}$. These two circles intersect the spiral at the point Q and Q'. The segments OQ and OQ' trisect the angle POA.

The curve of the nautilus shell is known as the *logarithmic spiral* or the *growth spiral*, and it was first discovered by Descartes. A logarithmic spiral is described by the equation

$$\log_b r = a\,\theta \tag{3}$$

which can also be expressed as

$$r = b^{a\,\theta}. \tag{4}$$

Figure 5 (over the page) shows the logarithmic spiral $\log_{1.2} = \theta$ produced with a spreadsheet program using the same technique as shown for the spiral of Archimedes. Use the same program to find out about the effect the parameters a and b have on the shape of the spiral. In particular, what happens if $a > 0$? What if $a < 0$? What if $a = 0$?

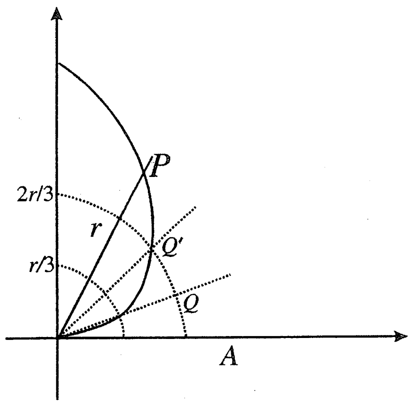


Figure 4

The loops of a logarithmic spiral are spaced farther and farther away; the distance between them follows a certain pattern. Given that the loops of the spiral of Archimedes are equidistant, equation (3) says that in the logarithmic spiral, the logarithms of successive radii are equally spaced.

So, if r_1, r_2, r_3, r_4, ... are the points where the logarithmic spiral cuts the horizontal axis, then

$$\log_b r_3 - \log_b r_2 = \log_b r_2 - \log_b r_1$$

which means that $\log_b \frac{r_3}{r_2} = \log_b \frac{r_2}{r_1}$ and therefore $\frac{r_3}{r_2} = \frac{r_2}{r_1}$. Hence the successive distances r_1, r_2, r_3, r_4, ... from the loop to the centre of the logarithmic spiral form a geometric sequence.

As a final exercise, I invite you to investigate a few other well known spirals: The *hyperbolic spiral* $r\,\theta = a$, where the radius is inversely proportional to the angle, *Fermat's spiral* (also know as the parabolic spiral) : $r^2 = a\,\theta$, *Fermat's spiral* (also know as the parabolic spiral) : $r^2 = a\,\theta$.

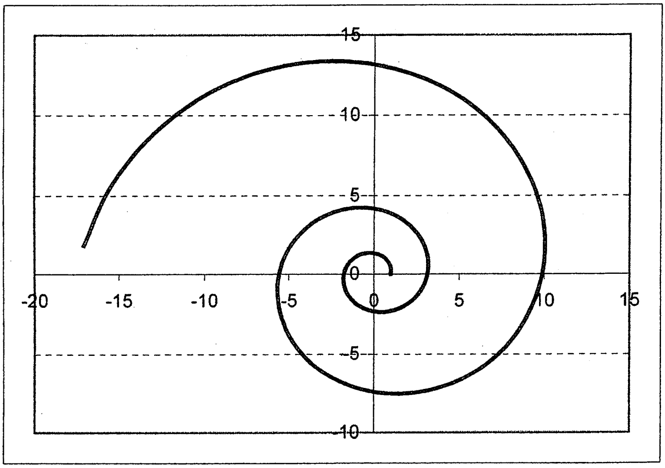


Figure 5

* * * * *

PROBLEMS AND SOLUTIONS

SOLUTION TO PROBLEM 25.3.1

This problem came from *Mathematical Digest*, a South African counterpart of *Function*. It read:

An adventurer who went in search of treasure on a certain small island had as sole clue the following instructions:

> From the middle of the hut (H) make a line to the ash tree (A) and another to the beech tree (B). From A in the direction remote from B make a line AC at right angles to and equal in length to HA. Similarly from B make a line BD. The treasure is buried at the midpoint of the line CD.

The adventurer arrived at the island and was able to identify the trees, but all trace of the hut had vanished owing to the ravages of termites. How was the treasure found?

Solutions were received from Keith Anker and Carlos Victor, and *Mathematical Digest* published yet another. What follows is a composite of these various solutions.

Put the origin of co-ordinates at the mid-point of AB and suppose that A is the point ($-a$, 0 and that B is the point (a, 0). Let the co-ordinates of the hut be (h, k). There will turn out to be three separate cases depending on whether k is positive, negative or zero. Consider first the case $k > 0$. Then HA and AC form two sides of a square and C has the co-ordinates ($-a-k$, $a+h$). Similarly D has the co-ordinates ($a+k$, $a-h$). The co-ordinates of the treasure are therefore (0, a). As we have taken $k > 0$, the situation is that, *when viewed from the hut*, A is to the right and B to the left. This corresponds to the case $a > 0$. Had we had $k < 0$, the situation would have been reversed, and in that case $a < 0$.

We thus have that the treasure is buried on the perpendicular bisector of AB and is either on one side or the other at a distance from the mid-point of the line-segment AB the same as the distance to either of the trees.

There are two points to investigate. They are given by $(\pm a, 0)$. The treasure will be at the one or the other.

It remains to consider the situation in which the hut is in line with the two trees, i.e. $k = 0$. The detail of this is left as an exercise for the reader. It is ambiguous as to how the word "remote" is to be interpreted in this case. Anker, in his solution, took the view that this rendered the case vacuous (and in any case it is unlikely on other grounds). However, the South African solvers tried to make sense of the instructions even when this situation applied and found a third possibility in which the treasure was to be found at the mid point of AB.

The solution may be "cooked", as the jargon has it, in the sense that one may find it by a "cheating" method. The argument runs as follows. Because the position of the hut is to make either no difference or very little difference, we may place it almost anywhere. A good choice is to put it on the mid-line between the trees, and this will simplify the argument already given. And other such choices are possible. The reader may care to explore these.

A similar "cooking" argument was noted in the case of Problem 25.2.1.

SOLUTION TO PROBLEM 25.3.2

This problem (submitted by Keith Anker) concerned a game played between two players. A set of nine tokens is arranged in a square as shown on the next page.

A move consists of the removal of any or all of the tokens in any one row or column. The players, A and B, move alternately with A going first, and the game continuing until all the tokens have been removed. The winner is the player taking the last of the tokens.

The problem asked which of the players can force a win, and what is the winning strategy?

$$\begin{array}{ccc} * & * & * \\ * & * & * \\ * & * & * \end{array}$$

Answers were received from Keith Anker (the proposer) and from Carlos Victor. Again we print a composite.

The answer is that B (the second player) can always force a win. To see this, consider first the similar but simpler game in which there are four tokens:

*	*
*	*

Here player A has two possible moves. One is to take both tokens from any row or column. (It is clear that it makes no difference which of the four such possibilities is actually adopted.) In that case, B wins immediately by taking all the remaining tokens. So A might resort to taking just one token. (Again it makes no difference which!) In this case, B takes the token diagonally opposite from the one that A has just taken, and it becomes clear that B now has a forced win.

Now consider the problem actually asked. Again it makes no difference which row or column we begin with. We may (actually or notionally) interchange the rows (or columns) with one another, and rows and columns interchange when we look at the pattern sideways. So begin by supposing that A takes *all* the tokens in any row (or column). In this case, B replies by taking all the remaining tokens in any column (or row). This leaves the 2×2 variant, which we have already analysed to a win for B.

A similar situation applies if A takes 2 tokens from any row or column. This necessarily leaves one column or row with 3 tokens, and B takes *that* to reach the same situation as detailed above. Thus A's only chance is to take a single token. To follow the analysis in this case, number the tokens as follows:

1	2	3
4	5	6
7	8	9

Because it makes no difference which row or column we adopt as our reference, it makes no difference which of the tokens *A* actually takes. The solution given below can be modified to cover all contingencies, merely by altering the labelling. Thus suppose that *A* takes token 3. *B* may then force the win by taking tokens 5 and 6.

For then, however *A* moves, *B* wins. With obvious notation:

(1, 4, 7) is answered by 8
(7, 8, 9) is answered by 1
(1, 2) is answered by (7, 9)
(1, 4) and (1, 7) are both answered by (8, 9)
(4, 7) is answered by (2, 8)
(7, 8) and (7, 9) are both answered by (1, 2)
1 is answered by (7, 8, 9)
2 is answered by 7 and vice versa (and an analysis akin to the 2×2 case)
4 is answered by 9 and vice versa
8 is answered by (1, 7).

SOLUTION TO PROBLEM 25.3.3

This problem (submitted by Garnet J Greenbury) read:

Let *ABC* be a triangle with its inscribed circle centred at *O* and one of its three escribed circles centred at *P*. The mid-point of the line-segment *OP* is *D*. Prove that *D* lies on the circumcircle of the triangle *ABC*.

The proposer submitted three different solutions, and we also received replies from Keith Anker, Julius Guest and Carlos Victor. We particularly liked Anker's solution, which went like this.

Suppose the excentres of the three escribed circles are *P*, *Q*, *R* opposite *A*, *B*, *C* respectively. Now the internal and the external bisectors of each of the angles *A*, *B*, *C* are at right angles to each other. Hence *O* is the orthocentre of the triangle *PQR*. Thus the circumcentre of the triangle *ABC* is on the 9-point circle of the triangle *PQR*. [For detail on the 9-point circle, see the cover article to *Function*, February 1979.] Another of the 9 points is the mid-point of *OP*.

SOLUTION TO PROBLEM 25.3.4

The problem asked for the final two digits of $(20n - 15)^m$, where m, n are positive integers. Solutions were received from Keith Anker, Julius Guest and Carlos Victor. Here is Victor's solution.

Put $k = n - 1$. If $m = 1$, $20n - 15 = 20k + 5$. So then the possible values of the final two digits are 05, 25, 35, 45, 65, 85.

If $m > 1$, then $(20k + 5)^m = 100r + 5^m$, where r is some positive integer. It follows that the final two digits must be 25.

Here is a further set of problems for readers to try.

PROBLEM 25.5.1 (submitted by Peter Grossman)

Given any four-digit number in which not all of the digits are the same, a sequence is generated as follows. The first term is the given number, and each term is used to determine the next term according to the following rule:

1. Rearrange the digits of the number into decreasing order, to obtain a four-digit number *ABCD*, and into increasing order, to obtain another four-digit number *DCBA* (where *A*, *B*, *C* and *D* denote the digits).
2. Subtract *DCBA* from *ABCD*.

(For example, starting with 5946, the sequence 5946, 5085, 7992, 7173, 6354, 3087, 8352, 6174, 6174, … is generated.)

Prove that the sequence must eventually reach 6174, regardless of the given number.

PROBLEM 25.5.2 (submitted by Peter Grossman)

Find all real number solutions in x and y of the equation $x^y = 1$.

PROBLEM 25.5.3 (submitted by Julius Guest)

Sum the series

$$S = \frac{2}{1!} + \frac{9}{2!} + \frac{28}{3!} + \ldots\ldots.. + \frac{n^3 - 1}{n!} + \ldots\ldots. \ .$$

PROBLEM 25.5.4 (from an article by Theodore Eisenberg in *International Journal of Mathematical Education in Science and Technology*)

Let a, b, c, d, e be digits in our usual decimal system. Prove that 7 divides the number $abcde$ ($7|\ abcde$) if and only if it also divides the number $(abcd - 2e)$, i.e. $7|\ (abcd - 2e)$.

****

The Drunken Walker and the Plastered Fly

Imagine a drunken person who starts wandering on the number line at 0, and then moves left or right (+/–1) with probability 1/2. What is the probability that the walker will eventually return to the starting point?

Answer: probability 1.

What about a random walk in the plane, moving on the integer lattice points, with probability 1/4 in each of the coordinate directions? What's the chance of return to the starting point?

Answer: probability 1.

But what about a plastered fly, with 6 directions to move, probability 1/6? Surprisingly, it is probable that the fly will never return to its start.

In fact it only has probability around 1/3 of ever returning. This is because there is so much "space" in dimensions 3 and higher.

Adapted from the Math Fun Facts website of Harvey Mudd College.

Index to Volume 25

Feature Articles

History Columns

Computers and Computing

Front Covers

Olympiad News

Miscellanea

Published by Department of Mathematics & Statistics, Monash University