4 Picture proofs

Do you ever walk through a proof, understand each step, yet not believe the theorem, not say 'Yes, of course it's true'? The analytic, logical, sequential approach often does not convince one as well as does a carefully crafted picture. This difference is no coincidence. The analytic, sequential portions of our brain evolved with our capacity for language, which is perhaps 10^5 years old. Our pictorial, Gestalt hardware results from millions of years of evolution of the visual system and cortex. In comparison to our visual hardware, our symbolic, sequential hardware is an ill-developed latecomer. Advertisers know that words alone do not convince you to waste money on their clients' junk, so they spend zillions on images. This principle, which has higher applications, is the theme of this chapter.

4.1 Adding odd numbers

Here again is the sum from **Section 2.1** that illustrated using extreme cases to find fencepost errors:

$$S = \underbrace{1 + 3 + 5 + \cdots}_{n \text{ terms}}$$

Before I show the promised picture proof, let's go through the standard method, proof by induction, to compare it later to the picture proof. An induction proof has three pieces:

1. Verify the *base case* $n = 1$. With $n = 1$ terms, the sum is $S = 1$, which equals n^2. QED (Latin for 'quite easily done').

2. Assume the *induction hypothesis.* Assume that the sum holds for n terms:

$$\sum_{1}^{n}(2k-1) = n^2.$$

This assumption is needed for the next step of verifying the sum for $n+1$ terms.

3. Do the *induction step* of verifying the sum for $n+1$ terms, which requires showing that

$$\sum_{1}^{n+1}(2k-1) = (n+1)^2.$$

The sum splits into a new term and the old sum:

$$\sum_{1}^{n+1}(2k-1) = \underbrace{2n+1}_{\text{new term}} + \sum_{1}^{n}(2k-1).$$

The sum on the right is n^2 courtesy of the induction hypothesis. So

$$\sum_{1}^{n+1}(2k-1) = 2n+1+n^2 = (n+1)^2.$$

The three parts of the induction proof are complete, and the theorem is proved. However, the parts may leave you feeling that you follow each step but do not see *why* the theorem is true.

Compare it against the picture proof. Each term in the sum S adds one odd number represented as the area of an L-shaped piece. Each piece extends the square by one unit on each side. Adding n terms means placing n pieces and making an $n \times n$ square. [Or is it an $(n-1) \times (n-1)$ square?] The sum is the area of the square, which is n^2. Once you understand this picture, you never forget why adding the first n odd numbers gives the perfect square n^2.
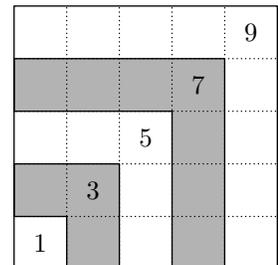
4.2 Geometric sums

Here is a familiar series:

$$S = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots.$$

The usual symbolic way to evaluate the sum is with the formula for a geometric series. You can derive the formula using a trick. First compute $2S$ by multiplying each term by 2:

$$2S = 2 + \underbrace{1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots}_{S}.$$

This sum looks like S, except for the first term 2. So $2S = 2 + S$ and $S = 2$.

The result, though correct, may seem like magic. Here then is a picture proof. A square with unit area represents the first term, which is $1/2^0$ (and is labelled 0). The second term is a $1 \times 1/2$ rectangle representing $1/2^1$ (and is labelled against by the exponent 1). The third term is a $1/2 \times 1/2$ square placed in the nook. The fourth term is, like the second term, a rectangle. With every pair of terms, the empty area between all the rectangles and three-quarters of the 1×2 outlining rectangle fills in. In the limit, the sum fills the 1×2 rectangle, showing that $S = 2$.
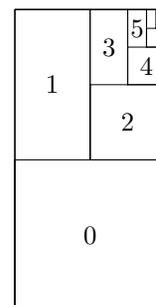
4.3 Arithmetic mean–geometric mean inequality

A classic inequality is the arithmetic mean–geometric mean inequality. Here are a few numerical examples before the formal statement. Take two numbers, say, 1 and 2. Their arithmetic mean is 1.5. Their geometric mean is $\sqrt{1 \times 2} = 1.414\ldots$. Now try the same operations with 2 and 3. Their arithmetic mean is 2.5, and their geometric mean is $\sqrt{2 \times 3} = 2.449\ldots$. In both cases, the geometric mean is smaller than the arithmetic mean. This pattern is the theorem of the arithmetic mean and geometric mean. It says that when $a, b \geq 0$, then

$$\underbrace{\frac{a+b}{2}}_{\text{AM}} \geq \underbrace{\sqrt{ab}}_{\text{GM}},$$

where AM means arithmetic mean and GM means geometric mean.

It has at least two proofs: symbolic and pictorial. A picture proof is hinted at by the designation of $\sqrt{ab}$ as the geometric mean. First, however, I prove it symbolically. Look at $(a - b)^2$. Since it is a square,

Picture proofs

$$(a - b)^2 \geq 0.$$

Expanding the left side gives $a^2 - 2ab + b^2 \geq 0$. Now do the magic step of adding $4ab$ to both sides to get

$$a^2 + 2ab + b^2 \geq 4ab.$$

The left side is again a perfect square, whose perfection suggests taking the square root of both sides to get

$$a + b \geq 2\sqrt{ab}.$$
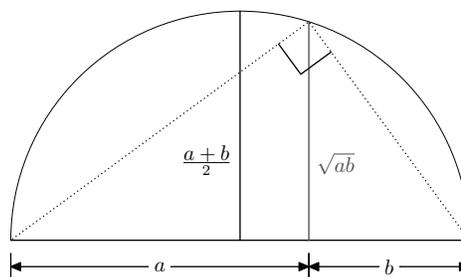
Dividing both sides by 2 gives the theorem:

$$\underbrace{\frac{a + b}{2}}_{\text{AM}} \geq \underbrace{\sqrt{ab}}_{\text{GM}}$$

Maybe you agree that, although each step is believable (and correct), the sequence of all of them seems like magic. The little steps do not reveal the structure of the argument, and the *why* is still elusive. For example, if the algebra steps had ended with

$$\frac{a + b}{4} \geq \sqrt{ab},$$

it would not have seemed obviously wrong. We would like a proof whose result *could not have been otherwise.*

Here then is a picture proof. Split the diameter of the circle into the lengths a and b. The radius is $(a + b)/2$, which is the arithmetic mean. Now we need to find the geometric mean, whose name is auspicious. Look at the second half chord rising from the diameter where a and b meet. It is also the height of the dotted triangle, and that triangle is a right triangle. With right triangles everywhere, similar triangles must come in handy. Let the so-far-unknown length be x. By similar triangles,

$$\frac{x}{a} = \frac{b}{x},$$

so $x = \sqrt{ab}$, showing that the half chord is the geometric mean. That half chord can never be greater than the radius, so the geometric mean is never greater than the arithmetic mean. For the two means to be equal, the geometric-mean half chord must slide left to become the radius, which happens only when $a = b$. So the arithmetic mean equals the geometric mean when $a = b$.

Compare this picture proof with the symbolic proof. The structure of the picture proof is there to see, so to speak. The only non-obvious step is showing that the half chord is the geometric mean $\sqrt{ab}$, the geometric mean. Furthermore, the picture shows why equality between the two means results only when $a = b$: Only then does the half chord become the radius.

Here are two applications of the AM–GM inequality to problems from introductory calculus that one would normally solve with derivatives. In the first problem, you get $l = 40\,\text{m}$ of fencing to mark off a rectangular garden. What dimensions does the garden have in order to have the largest area? If a is the length and b is the width, then $l = 2(a + b)$, which is $4 \times \text{AM}$. The area is ab, which is $(\text{GM})^2$. Since $\text{AM} \geq \text{GM}$, the consequence in terms of this problem's parameters is

$$\text{AM} = \frac{l}{4} \geq \sqrt{\text{area}} = \text{GM}.$$

Since the geometric mean cannot be larger than $l/4$, which is constant, the geometric mean is maximized when when $a = b$. For maximum area, therefore choose $a = b = 10\,\text{m}$ and get $A = 100\,\text{m}^2$.

The next example in this genre is a more difficult three-dimensional problem. Start with a unit square and cut out four identical corners, folding in the four edges to make an open-topped box. What size should the corners be to maximize the box volume? Call x the side length of the corner cutout. Each side of the box has length $1 - 2x$ and it has height x, so the volume is
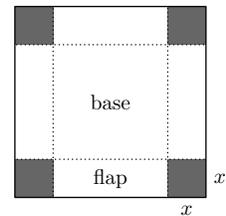
$$V = x(1 - 2x)^2.$$

For lack of imagination, let's try the same trick as in the previous problem. Two great mathematicians, George Polya and Gabor Szego, commented that, 'An idea which can be used once is a trick. If it can be used more than once it becomes a method.' So AM–GM, if it helps solve the next problem, gets promoted from a mere trick to the more exalted method.

In the previous problem, the factors in the area were a and b, and their sum $a + b$ was constant because it was fixed by the perimeter. Then we could use AM–GM to find the maximum area. Here, the factors of the volume are x, $1 - 2x$, and $1 - 2x$. Their sum is $2 - 3x$, which is not a constant; instead it varies as x changes. This variation means that we cannot apply the AM–GM theorem directly. The theorem is still valid but it does not tell us what we want to know. We want to know the largest possible volume. And, directly applied, the theorem says that the volume is never less than the cube of the arithmetic mean. Making the volume equal to this value does not guarantee that the maximum volume has been found, because the arithmetic mean is changing as one changes x to maximize the geometric mean. The largest volume may result where the GM is not equal to the changing AM. In the two-dimensional problem, this issue did not arise because the AM was already constant (it was a fixed fraction of the perimeter).

If only the factor of x were a $4x$, then the $3x$ would disappear when computing the AM:

$$4x + (1 - 2x) + (1 - 2x) = 2.$$

As Captain Jean-luc Picard of *The Next Generation* says, 'Make it so.' You can produce a $4x$ instead of an x by studying $4V$ instead of V:

$$4V = 4x \times 1 - 2x \times 1 - 2x.$$

The sum of the factors is 2 and their arithmetic mean is $2/3$ – which is constant. The geometric mean of the three factors is

$$(4x(1 - 2x)(1 - 2x))^{1/3} = (4V)^{1/3}.$$

So by the AM–GM theorem:

$$\mathrm{AM} = \frac{2}{3} \geq (4V)^{1/3} = \mathrm{GM},$$

so

$$V \leq \frac{1}{4}\left(\frac{2}{3}\right)^3 = \frac{2}{27}.$$

The volume equals this constant maximum value when the three factors $4x$, $1 - 2x$, and $1 - 2x$ are equal. This equality happens when $x = 1/6$, which is the size of the corner cutouts.

51

4.4   **Logarithms**

Pictures explain the early terms in many Taylor-series approximations. As an example, I derive the first two terms for $\ln(1+x)$. The logarithm function is defined as an integral

$$\ln(1+x) = \int_1^{1+x} \frac{dt}{t}.$$

An integral, especially a definite integral, suggests an area as its picture. As a first approximation, the logarithm is the area of the shaded, circumscribed rectangle. The rectangle, although it overestimates the integral, is easy to analyze: Its area is its width (which is x) times its height (which is 1). So the area is x. This area is the first pictorial approximation, and explains the first term in the Taylor series
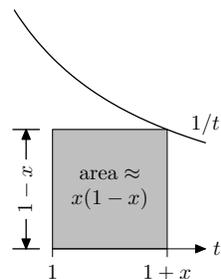
$$\ln(1+x) = x - \cdots.$$
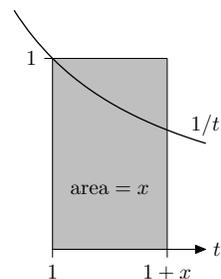
An alternative to overestimating the integral is to underestimate it using the inscribed rectangle. Its width is still x but its height is $1/(1+x)$. For small x,

$$\frac{1}{1+x} \approx 1 - x,$$

as you can check by multiplying both sides by $1+x$:

$$1 \approx 1 - x^2.$$

This approximation is valid when x^2 is small, which happens when x is small. Then the rectangle's height is $1 - x$ and its area is $x(1-x) = x - x^2$.

For the second approximation, average the over- and underestimate:

$$\ln(1+x) \equiv \text{area} \approx \frac{x + (x - x^2)}{2} = x - \frac{x^2}{2}.$$

These terms are the first two terms in the Taylor series for $\ln(1+ x)$. The picture for this symbolic average is a trapezoidal area, so this series of pictures explains the first two terms. Its error lies in making the smooth curve $1/t$ into a straight line, and this error produces the higher-order terms in the series – but they are difficult to compute just using pictures.

2008-03-06 13:24:47 / rev ebd336097912+
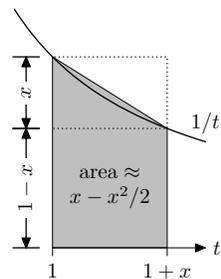
Picture proofs **52**

Alternatively you can derive all the terms from the binomial theorem and the definition of the logarithm. The logarithm is

$$\ln(1+x) \equiv \int_1^{1+x} \frac{dt}{t} = \int_0^x \frac{1}{1+t}\, dt.$$

The binomial theorem says that

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \cdots,$$

so

$$\ln(1+x) = \int_0^x \left(1 - t + t^2 - t^3 + \cdots\right) dt.$$

Now integrate term by term; although this procedure produces much gnashing of the teeth among mathematicians, it is usually valid. To paraphrase a motto of the Chicago police department, 'Integrate first, ask questions later.' Then

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots.$$

The term-by-term integration shows you the entire series. Understand both methods and you will not only remember the logarithm series but will also understand two useful techniques.

As an application of the logarithm approximation, I estimate $\ln 2$. A quick application of the first two terms of the series gives:

$$\ln(1+x) \approx x - \frac{x^2}{2}\Big|_{x=1} = 1 - \frac{1}{2} = \frac{1}{2}.$$

That approximation is lousy because x is 1, so squaring x does not help produce a small $x^2/2$ term. A trick, however, improves the accuracy. Rewrite $\ln 2$ as

$$\ln 2 = \ln \frac{4/3}{2/3} = \ln \frac{4}{3} - \ln 2/3.$$

Then approximate $\ln(4/3)$ as $\ln(1+x)$ with $x = 1/3$ and approximate $\ln(2/3)$ as $\ln(1+x)$ with $x = -1/3$. With $x = \pm 1/3$, squaring x produces a small number, so the error should shrink. Try it:

$$\ln \frac{4}{3} = \ln(1+x)\big|_{x=1/3} \approx \frac{1}{3} - \frac{1}{2} \cdot \left(\frac{1}{3}\right)^2,$$

$$\ln \frac{2}{3} = \ln(1+x)\big|_{x=-1/3} \approx -\frac{1}{3} - \frac{1}{2} \cdot \left(-\frac{1}{3}\right)^2.$$

2008-03-06 13:24:47 / rev ebd336097912+

When taking the difference, the quadratic terms cancel, so

$$\ln 2 = \ln \frac{4}{3} - \ln \frac{2}{3} \approx \frac{2}{3} = 0.666\ldots.$$
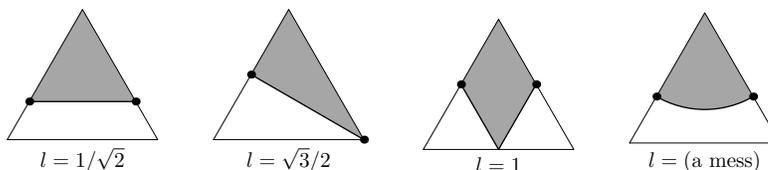
The true value is $0.697\ldots$, so this estimate is accurate to 5%!

4.5 Geometry

The following pictorial problem has a natural pictorial solution:

> How do you cut an equilateral triangle into two equal halves using the shortest, not-necessarily-straight path?

Here are several candidates among the infinite set of possibilities for the path.

$l = 1/\sqrt{2}$ $l = \sqrt{3}/2$ $l = 1$ $l = $ (a mess)

Let's compute the lengths of each bisecting path, with length measured in units of the triangle side. The first candidate encloses an equilateral triangle with one-half the area of the original triangle, so the sides of the smaller, shaded triangle are smaller by a factor of $\sqrt{2}$. Thus the path, being one of those sides, has length $1/\sqrt{2}$. In the second choice, the path is an altitude of the original triangle, which means its length is $\sqrt{3}/2$, so it is longer than the first candidate. The third candidate encloses a diamond made from two small equilateral triangles. Each small triangle has one-fourth the area of the original triangle with side length one, so each small triangle has side length $1/2$. The bisecting path is two sides of a small triangle, so its length is 1. This candidate is longer than the other two.

The fourth candidate is one-sixth of a circle. To find its length, find the radius r of the circle. One-sixth of the circle has one-half the area of the triangle, so

$$\underbrace{\pi r^2}_{A_{\text{circle}}} = 6 \times \frac{1}{2} A_{\text{triangle}} = 6 \times \frac{1}{2} \times \underbrace{\frac{1}{2} \times 1 \times \frac{\sqrt{3}}{2}}_{A_{\text{triangle}}}.$$

Multiplying the pieces gives

$$\pi r^2 = \frac{3\sqrt{3}}{4},$$

and

$$r = \sqrt{\frac{3\sqrt{3}}{4\pi}}.$$

The bisection path is one-sixth of a circle, so its length is

$$l = \frac{2\pi r}{6} = \frac{\pi}{3}\sqrt{\frac{3\sqrt{3}}{4\pi}} = \sqrt{\frac{\pi\sqrt{3}}{12}}.$$

The best previous candidate (the first picture) has length $1/\sqrt{2} = 0.707\ldots$. Does the mess of π and square roots produce a shorter path? Roll the drums...:

$$l = 0.67338\ldots,$$

which is less than $1/\sqrt{2}$. So the circular arc is the best bisection path *so far*. However, is it the best among all possible paths? The arc-length calculation for the circle is messy, and most other paths do not even have a closed form for their arc lengths.

Instead of making elaborate calculations, try a familiar method, symmetry, in combination with a picture. Replicate the triangle six times to make a hexagon, and also replicate the candidate path. Here is the result of replicating the first candidate (the bisection line going straight across). The original triangle becomes the large hexagon, and the enclosed half-triangle becomes a smaller hexagon having one-half the area of the large hexagon.
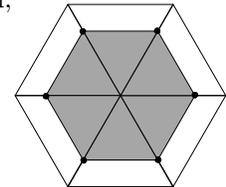
Compare that picture with the result of replicating the circular-arc bisection. The large hexagon is the same as for the last replication, but now the bisected area replicates into a circle. Which has the shorter perimeter, the shaded hexagon or this circle? The **isoperimetric theorem** says that, of all figures with the same area, the circle has the smallest perimeter. Since the circle and the smaller hexagon enclose the same area – which is three times the area of one triangle – the circle has a smaller perimeter than the hexagon, and has a smaller perimeter than the result of replicating any other path!
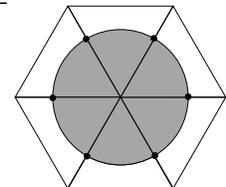
4.6   **Summing series**

Now let's look for a second time at Stirling's approximation to n factorial. In **Section 3.5**, we found it by approximating the integral

$$\int_0^\infty t^n e^{-t}\,dt = n!.$$

The next method is also indirect, by approximating $\ln n!$:

$$\ln n! = \sum_1^n \ln k.$$

This sum is the area of the rectangles. That area is roughly the area under the smooth curve $\ln k$. This area is

$$\int_1^n \ln k\,dk = k\ln k - k = n\ln n - n + 1.$$

Before making more accurate approximations, let's see how this one is doing by taking the exponential to recover $n!$:

$$n! \approx \frac{n^n}{e^n} \times e.$$

The n^n and the $1/e^n$ factors are already correct. The next pictorial correction make the result even more accurate.

The error in the integral approximation come from the pieces protruding beyond the $\ln k$ curve. To approximate the area of these protrusions, pretend that they are triangles. If $\ln k$ were made of linear segments, there would be no need to pretend; even so the pretense is only a tiny lie. The problem become one of adding up the shaded triangles.

The next step is to double the triangles, turning them into rectangles, and remembering to repay the factor of 2 before the end of the derivation.
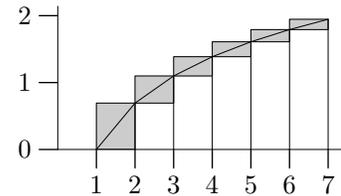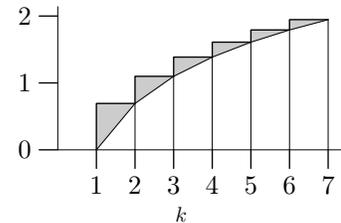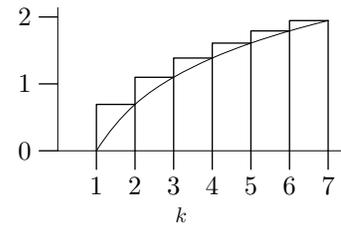
Picture proofs **56**

The final step is to hold your right hand at the $x = 7$ line to catch the shaded pieces as you shove them rightward with your left hand. They stack to make the $\ln 7$ rectangle. So the total overshoot, after paying back the factor of 2, is $(\ln 7)/2$. For general n, the overshoot is $(\ln n)/2$. The integral $\int_1^n \ln k \, dk$ provides $n \ln n - n$ (from the upper limit) and 1 from the lower limit. So the integral and graph together produce
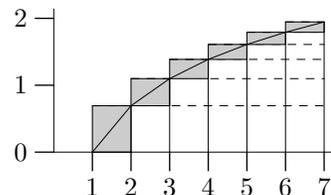
$$\ln n! \approx n \ln n - n + 1 + \underbrace{\frac{\ln n}{2}}_{\text{protrusions}}$$

or

$$n! \approx e\sqrt{n}\left(\frac{n}{e}\right)^n.$$

Stirling's formula is

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

The difference between the pictorial approximation and Stirling's formula is the factor of e that should be $\sqrt{2\pi}$. Except for this change of only 8%, a simple integration and graphical method produce the whole formula.

The protrusion correction turns out to be the first term in an infinite series of corrections. The later corrections are difficult to derive using pictures, just as the later terms in the Taylor series for $\ln(1 + x)$ are difficult to derive by pictures (we used integration and the binomial theorem for those terms). But another technique, analogy, produces the higher corrections for $\ln n!$. That analysis is the subject of **Section 7.3**, where the pictorial, protrusion correction that we just derived turns out to be the zeroth-derivative term in the Euler–MacLaurin summation formula.