

Pi

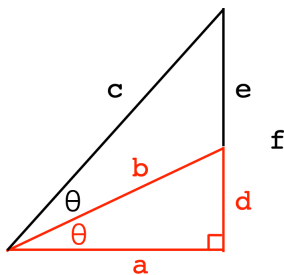
In this chapter we explore the math relating to Archimedes' method to approximate π , resulting in the bounds $3 \frac{10}{71} < \pi < 3 \frac{1}{7}$. We don't recapitulate his precise calculations, for that you can see

<https://itech.fgcu.edu/faculty/clindsey/mhf4404/archimedes/archimedes.html>

or another chapter in which I've interpreted that page.

As you probably know, the method involves calculating the length of the perimeter of circumscribed and inscribed regular polygons. We'll connect his formulas to those of Gregory relating to perimeter and area of such polygons.

We can build on two basic results from the geometry of right triangles:



The first is that the sum of the cotangent and cosecant for the original angle are equal to the cotangent of the half-angle:

$$\frac{a}{f} + \frac{c}{f} = \frac{a}{d}$$

$$\cot 2\theta + \csc 2\theta = \cot \theta$$

This follows from the angle bisector theorem, $a/d = c/e$, with a bit of manipulation.

Another way to get it is to use the double angle formulas. We'll abbreviate \sin, \cos, \tan as S, C, T and identify the values for the half-angle with a prime. The formulas are:

$$\begin{aligned} S &= 2S'C' \\ C &= C'^2 - S'^2 \\ &= 2C'^2 - 1 \end{aligned}$$

The cotangent of 2θ is

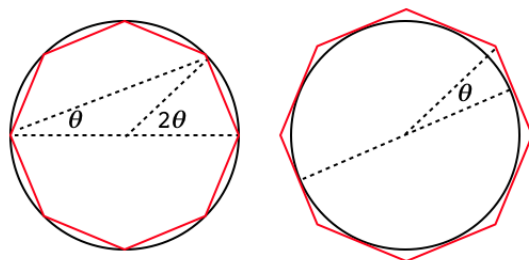
$$\begin{aligned} \frac{1}{T} &= \frac{2C'^2 - 1}{2S'C'} \\ &= \frac{1}{T'} - \frac{1}{S} \end{aligned}$$

which rearranges to give the result

$$\frac{1}{T'} = \frac{1}{T} + \frac{1}{S}$$

circumscribed polygon

Draw a right triangle, one side of which is part of a polygon that circumscribes a circle, with a central angle of 30° (right panel). (This figure from the web shows an octagon rather than the hexagon that Archimedes used, but it's the same idea).



We will estimate π as the ratio of the sum of the side lengths (the perimeter) of the polygon, to the diameter, since for a circle $\pi d = C$.

For a 30° right triangle, the sides are in the ratio $1, \sqrt{3}, 2$. The cotangent of θ is $\sqrt{3}$ and the cosecant is 2.

The ratio of a half-side to the radius is the tangent $1/\sqrt{3}$ and the ratio of a whole side to the diameter is exactly the same. Since a whole side corresponds to an angle of $2\theta = 60^\circ$, there are six such sides.

We estimate:

$$\pi < 6 \cdot \frac{1}{\sqrt{3}} \approx 3.4641$$

algorithm

To improve the estimate, obtain the cotangent of the half-angle.

The cosecant of $\theta = 30^\circ$ is 2 so, by addition, $\cot \theta/2$ is $2 + \sqrt{3}$. Multiply the inverse by the number of sides, which is now 12:

```
>>> sqrt3 = 3**0.5
>>> 1/(2 + sqrt3) * 12
3.2154...
```

For a right triangle with sides a, b, h Pythagoras says that

$$\begin{aligned} a^2 + b^2 &= h^2 \\ \frac{a^2}{b^2} + 1 &= \frac{h^2}{b^2} \end{aligned}$$

If the cotangent is a/b , then the cosecant is h/b , and the above formula can be used to convert the cotangent to the cosecant. With both cotangent and cosecant in hand, we are ready for another round.

Here's the first one:

$$\csc \theta/2 = \sqrt{\left(\frac{1}{2 + 265/153}\right)^2 + 1}$$

I wrote a short script to do the calculations.

<https://gist.github.com/telliott99/2c78eb20769e623b69c2c4c69adefa7b>

The output of successive approximations is:

```
3.464102
3.215390
3.159660
3.146086
3.142715
```

Going as far as Archimedes did, we diverge only in the thousandths place. That is four doublings starting with a hexagon: 12, 24, 48 and finally 96 sides.

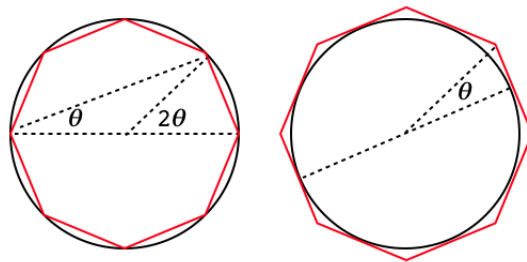
If you want to see exactly how he did all the calculations, you can look at my source

<https://itech.fgcu.edu/faculty/clindsey/mhf4404/archimedes/archimedes.html>

or you can read the chapter where I go through his argument in more detail, in a book on Analytic Geometry.

inscribed polygon

Draw a right triangle to be inscribed in a circle, with a central angle of 60° (left panel).



The geometry is similar to before, but with two differences. The ratio we seek is that of the opposing side to the hypotenuse (i.e. the sine). So we do the calculations exactly as before but multiply the inverse of the cosecant by the number of sides to get the circumference.

Also, the ratio is already that of a full side to the diameter.

The initial estimate is

$$\pi > 6 \cdot \frac{1}{2} = 3.0$$

The cotangent of $\theta/2$ is $2 + \sqrt{3}$ as we saw above. Go through the square plus one, then square root to get the cosecant, and then multiply by 12.

I modified the script to print the number of sides times the sine at each step:

3.464102	3.000000
3.215390	3.105829
3.159660	3.132629
3.146086	3.139350
3.142715	3.141032

Again, we diverge very little from Archimedes' values.

His final result was

$$3 \frac{10}{71} < \pi < 3 \frac{1}{7}$$

$$\frac{223}{71} < \pi < \frac{22}{7}$$

```
>>> pi
3.141592653589793
>>> 223/71.0
3.140845070422535
>>> 22/7.0
3.142857142857143
```

A rational approximation to π that was used by Ptolemy from the 2nd century C.E. is $377/120$, but an even better one is $355/113$.

```
>>> pi
3.141592653589793
>>> 377/120.0
3.1416666666666666
>>> 355/113.0
3.1415929203539825
```

The latter is due to the later (5th-century) Chinese mathematician Zu Chongzhi.

square root of 3

Archimedes needs values for the square root of 3 to do these calculations. Let's remember that, like π , $\sqrt{3}$ is an irrational number. It does not have a rational representation. That is to say, there are no whole numbers a and b such that

$$\frac{a^2}{b^2} = 3$$
$$a^2 = 3b^2$$

The proof for the irrationality of $\sqrt{3}$ is by contradiction:

Proof.

Assume that there are two integers a and b such that $a^2 = 3b^2$.

Use the *Fundamental theorem of arithmetic*, which says that every number has a unique prime factorization.

For example, if a is prime, then $a = p_1$, but otherwise $a = p_1 \cdot p_2 \dots$, and similarly $b = q_1 \cdot q_2 \dots$. Then a^2 will have an even number of prime factors, since there are two copies of each one, and similarly for b^2 .

On the left-hand side of the equation $a^2 = 3b^2$ then, by the theorem, there is an even number of factors. Since 3 is prime and b^2 is even, on the right-hand side there is an odd number of factors.

a^2 is an integer and it has a unique prime factorization, which means that this imbalance of factors is impossible. Hence there is no such pair of integers a and b .

This is a proof that the square root of every prime number is irrational.

□

As an ancient Greek, Archimedes uses ratios of whole numbers. Here are two excellent approximations:

$$\frac{265}{153} < \sqrt{3} < \frac{1351}{780}$$

It isn't clear how he came up with these. One possibility is that he used formulas that take an approximation to a square root and improve it. I found some discussion here:

<https://math.stackexchange.com/questions/894862/archimedes-approximation-of-square-root-1740660>

and here

https://homepage.divms.uiowa.edu/~atkinson/ftp/ENA_Materials/Overheads/sec_3-3.pdf

Newton-Raphson, or the Babylonian method

Suppose we have a first approximation to \sqrt{N} , call it x . We wish to find a better guess g as to the value of \sqrt{N} . If we compute

$$g = \frac{N}{x}$$

g and x "straddle" the true value.

[*Proof.* Suppose that $x^2 < N$. Then $1/x^2 > 1/N$, $N^2/x^2 > N$, thus $g^2 > N$. The converse is true also].

Furthermore, g is closer, it's a better guess.

The geometric mean of g and x is

$$\sqrt{g \cdot x} = \sqrt{g \cdot \frac{N}{g}} = \sqrt{N}$$

which would give us the precise value, but of course that assumes exactly what we're trying to find.

The arithmetic mean might do.

$$g = \frac{1}{2} (x + N/x)$$

The source calls this the "Babylonian method", but I've always known it as the Newton-Raphson method. It's a linear approximation. Technically, the Newton method applies to (almost) any $f(x)$ and is written in terms of $f'(x)$, while the Babylonian method is strictly for the square root function.

A simplified derivation is as follows. We can formulate the square root problem as $y = x^2 - N$ where we want to find the positive root x such that $y = 0 = x^2 - N$ since then $x^2 = N$.

For this parabola, the slope of the tangent line at any point such as $(x, x^2 - N)$ is $2x$ (from basic calculus or inspired geometry).

Let the zero of the tangent line be at the point $(g, 0)$, then we can write the point-slope equation of the tangent line as

$$\begin{aligned} 2x &= \frac{(x^2 - N) - 0}{x - g} \\ x - g &= \frac{x^2 - N}{2x} = \frac{1}{2} \left(x - \frac{N}{x} \right) \\ g &= \frac{1}{2} \left(x + \frac{N}{x} \right) \end{aligned}$$

Use the tangent line as an approximation to the parabola, and the point where the tangent line crosses the y -axis as an approximation to the zero of the parabola, that is, to \sqrt{N} .

As an example, $7/4$ is a reasonable first approximation of $\sqrt{3}$. Then

$$x = \frac{1}{2} \left(\frac{7}{4} + \frac{4}{7} \cdot 3 \right) = \frac{1}{2} \left(\frac{49 + 48}{28} \right) = \frac{97}{56}$$

A more sophisticated derivation is from here:

<http://www.math.ubc.ca/~anstee/math104/104newtonmethod.pdf>

Let r be the actual value of the zero of $f(x)$. Let x_0 be a good estimate of r , and the difference $h = r - x_0$. Linear approximation gives

$$f(r) = f(x_0 + h) \approx f(x_0) + f'(x_0) \cdot h$$

Starting from the value of the function at x_0 , a small change h gives a change in the value of the function of the derivative times the small change h .

And then (provided $f'(x_0)$ is not near zero):

$$f(r) = 0 \approx f(x_0) + f'(x_0) \cdot h$$

$$h \approx -\frac{f(x_0)}{f'(x_0)}$$

so

$$r = x_0 + h \approx x_0 - \frac{f(x_0)}{f'(x_0)}$$

In the case of the square root problem, the numerator is $x_0^2 - N$ and $f'(x_0) = 2x_0$ so

$$r \approx x_0 - \frac{x_0^2 - N}{2x_0}$$

Let us call the new value x_1

$$\begin{aligned} x_1 &= x_0 - \frac{x_0^2 - N}{2x_0} \\ &= \frac{1}{2} \left(x_0 + \frac{N}{x_0} \right) \end{aligned}$$

secant method

There is another method called the *secant* method. In Newton's method, we need the derivative. The secant method approximates the derivative by using the slope connecting two points on the curve.

Recall that $f(x) = x^2 - N$. Suppose we have two approximations to the square root, call them x_1 and x_2 . For these two points on the curve we have that the slope of the secant line connecting them is

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} = \frac{(x_2^2 - N) - (x_1^2 - N)}{x_2 - x_1} \\ &= \frac{x_2^2 - x_1^2}{x_2 - x_1} \end{aligned}$$

Since the numerator is a difference of squares, we see that m is just equal to $x_1 + x_2$.

Now let g also be on the same line, at the point where the line goes through the x -axis and $y = 0$. The point-slope equation (again) is

$$\begin{aligned} m = x_1 + x_2 &= \frac{f(x_1) - 0}{x_1 - g} = \frac{x_1^2 - N}{x_1 - g} \\ x_1 - g &= \frac{x_1^2 - N}{x_1 + x_2} \\ g &= \frac{N - x_1^2}{x_1 + x_2} + x_1 \\ &= \frac{N - x_1^2 + x_1^2 + x_1x_2}{x_1 + x_2} \\ &= \frac{N + x_1x_2}{x_1 + x_2} \end{aligned}$$

As an example, suppose we use $5/3$ and $7/4$ as approximations to $\sqrt{3}$. Then

$$g = \frac{3 + (5/3 \cdot 7/4)}{5/3 + 7/4}$$

It's a fraction of fractions, but the denominators cancel. So

$$g = \frac{36 + 5 \cdot 7}{4 \cdot 5 + 3 \cdot 7} = \frac{71}{41}$$

The last reference above also discusses the secant method, its connection to the Newton method, and also something about what Newton actually did.

brute-force search

To investigate further, I wrote a script

<https://gist.github.com/telliott99/09fee3aca365a91d9ca6019567b43b44>

that goes through steps through increasing values for the denominator and for each one, finds values for the numerator that give the closest approximation. For a given d , start with n so that $n/d < \sqrt{3}$ and then increment n until the ratio $n/d > \sqrt{3}$, by comparing $3 \cdot d^2$ to n^2 .

This gives two ratios $n/d > \sqrt{3}$ and $n - 1/d < \sqrt{3}$. Next, test if one of the two is very close to exactly right (we know it can never be precisely correct).

After filtering examples where neither ratio is close, the script output is:

```
low : 2    5    3
high: 1    7    4
high: 6    9    5
low : 3   12    7
high: 4   14    8
low : 2   19   11
high: 1   26   15
high: 6   33   19
low : 3   45   26
high: 4   52   30
low : 2   71   41
high: 1   97   56
high: 6  123   71
low : 3  168   97
high: 4  194  112
low : 2  265  153
high: 1  362  209
high: 6  459  265
low : 3  627  362
high: 4  724  418
```

The first number is how close we are. For example, with $7/4$, $7^2 = 49$ which is 1 more than $3 \cdot 4^2 = 48$. If you go down the list you will find $265/153$. However, I

wonder whether Archimedes computed squares of the first several hundred integers to find ratios this way. It's not elegant enough.

The list contains small ratios, that might easily be found by hand, and we can plug them into the approximation functions above. To do that I wrote another script

<https://gist.github.com/telliott99/efd23d51d51f5105acc1d3bdbb9eca45>

The output of the first part is

```
5/3 26/15
7/4 97/56
12/7 97/56
7/4 97/56
19/11 362/209
26/15 1351/780
45/26 1351/780
26/15 1351/780
71/41 5042/2911
97/56 18817/10864
168/97 18817/10864
```

What this says is that if we run the Newton-Raphson algorithm described above on the input in the first column, we obtain the result in the second. For example, starting with the ratio 26/15, we obtain the second of Archimedes' approximations, 1351/780. In fact, several other inputs give the same output, which is curious.

However, I do not find 265/153.

The second part of the script above uses the secant method with pairs of smaller ratios in the first list. The output is:

```
7/4 5/3 71/41
12/7 7/4 168/97
7/4 12/7 168/97
19/11 7/4 265/153
26/15 19/11 989/571
45/26 26/15 2340/1351
26/15 45/26 2340/1351
71/41 26/15 3691/2131
97/56 71/41 13775/7953
168/97 97/56 32592/18817
```

If you look at the output you will see that the second method (the secant method)

takes inputs $7/4$ and $19/11$ and finds $265/153$. This only worked by accident! It happened because I only tested adjacent values as inputs, and the value $14/8$ (not lowest terms) was right before $19/11$. (The output is much more voluminous if we do all pairwise tests, so I didn't show it here).

In any event, this is a reasonable idea about how Archimedes did it, with tools that might have been familiar to him.

It may also be possible that he used continued fractions (see the Analytic Geometry book). Once you understand the trick, that method is very easy, and there is a connection to Euclid's algorithm for the greatest common divisor, but we just don't know.

Here is the continued fraction representation of $\sqrt{3}$:

$$\sqrt{3} = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{1 + \frac{1}{2 + \dots}}}}$$

We obtain an approximation of the true value by ignoring the dots. So the last fraction is $1/2 + 1 = 3/2$. Continually, invert and add alternating 1 or 2, but finish with two 1's:

$$1 + 1/2 = 3/2$$

$$2 + 2/3 = 8/3$$

$$1 + 3/8 = 11/8$$

$$2 + 8/11 = 30/11$$

$$1 + 11/30 = 41/30$$

$$2 + 30/41 = 112/41$$

$$1 + 41/112 = 153/112$$

$$1 + 112/153 = 265/153$$

Want to know the trick? Write

$$(\sqrt{3} - 1)(\sqrt{3} + 1) = 2$$

So a first rearrangement

$$\sqrt{3} - 1 = \frac{2}{\sqrt{3} + 1} = \frac{2}{2 + \sqrt{3} - 1} = \frac{1}{1 + \frac{\sqrt{3}-1}{2}}$$

and then a second one

$$\frac{\sqrt{3}-1}{2} = \frac{1}{\sqrt{3}+1}$$

Substituting the second rearrangement into the first:

$$= \frac{1}{1 + \frac{1}{\sqrt{3}+1}} = \frac{1}{1 + \frac{1}{2+\sqrt{3}-1}}$$

Substitute from the first rearrangement:

$$= \frac{1}{2 + \frac{1}{1 + \frac{\sqrt{3}-1}{2}}}$$

and then the second

$$= \frac{1}{2 + \frac{1}{1 + \frac{1}{\sqrt{3}+1}}}$$

We are back where we started. So now both terms repeat:

$$\sqrt{3} = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{1 + \frac{1}{2 + \dots}}}}$$

which is written as $[1 : (1, 2)]$, and can be turned into an approximate fraction as shown above.

perimeter formula

There are other formulas to approximate π , two based on perimeters and two based on areas. Let's start with perimeters (circumference).

If you look above, you'll see that the perimeter of the two polygons is

$$P = nT$$

$$p = nS$$

(where capital letters are for the circumscribed figure, and lowercase for the inscribed one). As before, we will use primes to indicate the new values for the half-angle. It's important to remember that going to the new values, there is a factor of two from the side-doubling.

Thus we could write:

$$\begin{aligned} P' &= n'T' = 2nT' \\ p' &= n'S' = 2nS' \end{aligned}$$

Recall the formula relating S and T to T' , namely:

$$\frac{1}{T'} = \frac{1}{T} + \frac{1}{S}$$

The factors of n cancel, so in substituting we just get:

$$\begin{aligned} \frac{2}{P'} &= \frac{1}{P} + \frac{1}{p} \\ \frac{1}{P'} &= \frac{1}{2} \left(\frac{1}{P} + \frac{1}{p} \right) \end{aligned}$$

The new value P' is the *harmonic mean* of P and p .

We also need p' . Use the double-angle formula for sine:

$$\begin{aligned} S &= 2S'C' \\ S' &= \frac{S}{2C'} \\ &= \frac{S}{2} \frac{T'}{S'} \end{aligned}$$

The factors of n cancel, so in substituting we get:

$$\begin{aligned} \frac{p'}{2} &= \frac{p}{2} \frac{P'}{p'} \\ p'^2 &= pP' \\ p' &= \sqrt{pP'} \end{aligned}$$

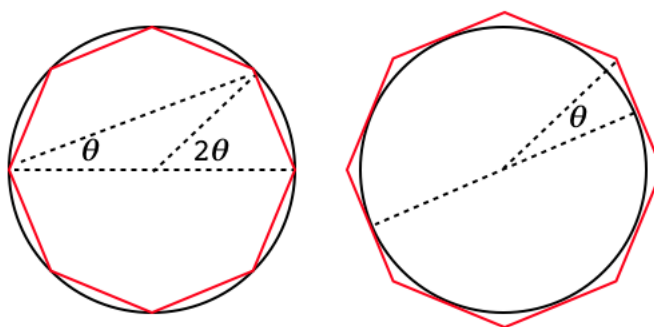
Suppose we start with p and P . We must find P' using the first formula before using the second one to find p' .

If you look at these the right way, you will see that we're doing the same operations as with the cotangent and the cosecant.

area formula

There is yet another way to apply the method, and that is to calculate the *areas* of inscribed and circumscribed polygons. Let's go through this briefly here and look more carefully at the geometry in the Analytic Geometry book.

For this approach we use a **unit circle** (radius 1) rather than a diameter of 1, as we did above. As before, we define θ to be the central angle of the half-sector (i.e. $\theta = 2\pi/2n$).



Rather than draw an entirely new figure, just imagine in the left panel that we draw the angle bisector of the central angle 2θ .

Each half triangle has base $\cos \theta$ and height $\sin \theta$, but we combine them, so the total area of the whole triangle is just $\sin \theta \cos \theta$ and the total area of the inner polygon is

$$a = n \sin \theta \cos \theta = nSC$$

in the notation of this chapter.

As before, to progress to a' we have a factor of 2 as well as the new values S' and C' :

$$a' = 2nS'C'$$

For the circumscribed or outer polygon, we just have what we had before, that the side length of the triangle in the right panel is $\tan \theta$ so the total area is

$$A = nT$$

Bring in the half-angle formulas as follows:

$$a' = 2nS'C' = 2n \cdot \frac{S}{2C'} \cdot C' = nS$$

That is slick, but we need an expression for nS in terms of the areas:

$$aA = nSC \cdot n\frac{S}{C} = [nS]^2$$

$$aA = [a']^2$$

$$a' = \sqrt{aA}$$

This is like, and yet subtly different than what we had when calculating the perimeter.

Since

$$A = nT$$

and

$$\begin{aligned} A' &= 2nT' \\ &= 2n \frac{ST}{S+T} = 2 \frac{nS \cdot nT}{nS + nT} \end{aligned}$$

$$A' = 2 \frac{a'A}{a' + A}$$

Compare

$$a' = \sqrt{aA} \quad A' = 2 \frac{a'A}{a' + A}$$

$$p' = \sqrt{pP'} \quad P' = 2 \frac{pP}{p + P}$$

We have primed values in corresponding positions.

However, it turns out that when you take account of the differing size of the circle for perimeter and area methods, and thus the initial values of p , P , a and A , the different order of operations results in precisely the same calculation.

I have transcribed a geometric derivation of the area formulas, which is due to Gregory. It's a very beautiful proof. This, plus the calculations, plus Archimedes exact ratios, are all in my book.