From Wallis’ formula to the Gaussian distribution and beyond

Mark Reeder
Department of Mathematics, Boston College
Chestnut Hill, MA 02467

April 25, 2015

Contents

1 The Wallis formula for \( \pi \) .................................................. 2

2 The basic integrals .................................................................. 7

2.1 Calculating the integrals \( I_n \) .............................................. 9

3 Wallis and coin-tossing .......................................................... 13

4 The Gaussian integral and factorial of one-half ..................... 17

5 Volumes of spheres ................................................................ 23

6 Wallis and the binomial distribution .................................. 26

7 The function \( \text{erf}(x) \). .......................................................... 29

8 From Wallis to Newton: The Binomial Theorem .................. 35

9 Binomial Theorem and the ellipse ...................................... 38
1 The Wallis formula for $\pi$

And he made a molten sea, ten cubits from the one brim to the other: it was round all about, and his height was five cubits: and a line of thirty cubits did compass it round about. (I Kings 7, v.23).

This ancient text says that $\pi = 3.0$. Actually, much better approximations were known prior to this, for example, the Egyptians used the approximation 3.16. Now, of course, billions of digits of $\pi$ are known. The next 57 of them are

$$\pi = 3.14159265358979323846264338327950288419716939937510582097\ldots$$

These digits do not repeat themselves, and have no recognized pattern. However, John Wallis (1616-1703, Savilian professor of Geometry at Oxford) discovered that there is a pattern, if we write $\pi$ as a product of fractions, instead of a sum of powers of 10. Wallis found that

$$\frac{\pi}{2} = \frac{2 \cdot 4 \cdot 6 \cdot 8 \cdot 
\ldots}{1 \cdot 3 \cdot 5 \cdot 7 \ldots}$$

(1)

Note that we have written the right side as a product of little fractions; we have not written

$$\frac{\pi}{2} = \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \ldots}{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \ldots}$$

because we might be tempted to cancel the odds below by their doubles above, and get

$$\frac{\pi}{2} = \frac{2 \cdot 2 \cdot 4 \cdot 2 \cdot 8 \cdot 8 \cdot 2 \cdot 2 \ldots}{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \ldots}$$

which is absurd.

To be clear, let us define the **Wallis fractions**:

$$W_k = \frac{2 \cdot 4 \cdot 6 \cdot 8 \ldots}{1 \cdot 3 \cdot 5 \cdot 7 \ldots} = \frac{2k}{2k-1}$$

so

$$W_1 = \frac{2 \cdot 2}{1 \cdot 3} = \frac{4}{3}, \quad W_2 = \frac{2 \cdot 4}{1 \cdot 3 \cdot 5} = \frac{64}{45}$$

etc.

Then the precise statement of Wallis' formula is

$$\lim_{k \to \infty} W_k = \frac{\pi}{2}$$
Let us try to convince ourselves of this, using a computer. We’ll actually compute $2 \cdot W_k$, which approaches the more recognizable decimal $\pi$. We find that

\[
2 \cdot W_3 = 2.92571\ldots, \quad 2 \cdot W_4 = 2.97215\ldots, \quad W_5 = 3.00218\ldots, \quad 2 \cdot W_{100} = 3.13379\ldots, \quad 2 \cdot W_{1000} = 3.14081\ldots
\]

So it seems to be getting there, but very slowly. When we say “limit as $k$ goes to infinity” we mean it! The above approximations are so weak that it seems no one armed with just a quill pen and expensive parchment would be able to guess that the sequence $2W_k$ actually converges to $\pi$. Wallis arrived at his formula for $\pi$ by a wild and creative path, guided by guessing and intuition, along with lots of persistance. It appears as Proposition 191 in his book *Arithmetica Infinitorum* (Arithmetic of Infinitesimals) published in 1656. This book was Wallis’ effort to find systematic methods to compute areas under curves, based on recent new ideas of Cavallieri, as expounded by Torricelli. In those days, finding area was called “quadrature” from the Latin *quadratus*, meaning “square”. The aim was to express the area of curved regions in terms of square units. In Wallis’ day, the most famous quadrature was Archimedes’ *Quadrature of the Parabola*, which set the standard. In his preface to *Arithmetica Infinitorum*: Wallis writes:

And indeed if the quadrature of one parabola rendered so much fame to Archimedes (so that then all mathematicians since that time placed him as though on the columns of Hercules), I felt it would be welcome enough to the mathematical world if I taught the quadrature also of infinitely many kinds of figures of this sort.

When Wallis showed his formula to the Christian Huygens (1629-1695, inventor of the pendulum clock), the latter was highly skeptical until Wallis could demonstrate that the right side of (1) agreed with $\pi/2$ to at least nine decimal places. Although $W_{1000}$ is nowhere close to this, William Brouckner (1620-1684, first President of the Royal Society) showed Wallis a remarkable way approximate Wallis’ product using continued fractions. This occupies the mystifying last few pages of *Arithmetica Infinitorum*, and we’ll eventually understand this via the much later work of Thomas Stieltjes (1856-1894).

Let us see how Wallis found his formula. First of all, the letter $\pi$, standing for the “periphery” of a circle with unit diameter, seems to have been first used by the Welsh mathematician William Jones in his textbook, *Synopsis Palmariorum Mathesos*, printed in 1706 just after Wallis’ death in 1703. So Wallis had his own private notation; he wrote $\Box$ to stand for what we call $\frac{4}{\pi}$. Thus,

---

1. After three and a half centuries, the *Arithmetica Infinitorum* has finally been translated into English, by Jacqueline Stedall: *The Arithmetic of Infinitesimals*, Springer-Verlag, 2004. See also [commentaries]
Wallis wrote

\[
\frac{1}{2\pi} = \int_0^1 (1 - x^2)^{1/2} dx = \frac{\pi}{4}.
\] (2)

The goal is to compute the integral explicitly, to get a formula for \(\pi\). Wallis does not have the Binomial Theorem (that must wait for Newton) so he cannot expand the integrand to integrate term by term. However, if he switched the 2 and the \(1/2\), he could expand:

\[
\int_0^1 (1 - x^{1/2})^2 dx = \int_0^1 1 - 2x^{1/2} + x \, dx = 1 - 2 \cdot \frac{2}{3} + \frac{1}{2} = \frac{1}{6}.
\]

Wallis sees that for any integers \(p, q\), he can similarly compute the integral

\[
\int_0^1 (1 - x^{1/p})^q dx
\] (3)

by expanding the integrand. He does this for various \(p, q\) and arrives at the first table below. The integral (3) is always 1 divided by an integer, and he just writes the integer, so the entry in row \(p\) and column \(q\) in the table below is

\[
\left[ \int_0^1 (1 - x^{1/p})^q \, dx \right]^{-1}.
\] (4)

<table>
<thead>
<tr>
<th>(p)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>21</td>
<td>28</td>
<td>36</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>20</td>
<td>35</td>
<td>56</td>
<td>84</td>
<td>120</td>
<td>165</td>
<td>220</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5</td>
<td>15</td>
<td>35</td>
<td>70</td>
<td>126</td>
<td>210</td>
<td>330</td>
<td>495</td>
<td>715</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6</td>
<td>21</td>
<td>56</td>
<td>126</td>
<td>252</td>
<td>462</td>
<td>792</td>
<td>1287</td>
<td>2002</td>
</tr>
</tbody>
</table>

Actually, Wallis computed many more entries than this, because, by examining this table, he hoped to discover a general formula, in terms of \(p\) and \(q\), for the integral (3). Then, if this formula still held true even if \(p\) and \(q\) were not integers, he could substitute \(p = q = 1/2\) (since those are the numbers in his original integral (2)) and he would have a formula for \(\Box\), hence a formula for \(\pi\)! However, things did not go so smoothly. Wallis writes:

Although no small hope seemed to shine, what we have in hand is slippery, like Proteus, who in the same way, often escaped, and disappointed hope.

\(^2\textit{Arithmetica Infinitorum},\) note 16 to Proposition 189
You may have already noticed the Binomial Triangle in Wallis’ table: The entry in row $p$, column $q$ is

$$a_{p,q} = \frac{(p + q)!}{p! \cdot q!} = \left[ \int_0^1 (1 - x^{1/p})^q \, dx \right]^{-1}.$$  \hspace{1cm} (5)

For example, the numbers in the second row are

$$a_{2,q} = (1/2)(q + 1)(q + 2),$$

and in the third row we have

$$a_{3,q} = \frac{1}{5}(q + 1)(q + 2)(q + 3),$$

and so on. But remember that Wallis wants $p$ and $q$ to be half-integers. The above formulas for $a_{p,q}$ make sense for any number $q$. So Wallis attempts the method of interpolation, where you look for a formula that is known to hold for integers, and whose terms make sense for all numbers. Then you hope that the formula holds for for all numbers.

The trouble is that Wallis wants to compute $a_{(1/2),(1/2)}$, but the formula (5) gives

$$a_{(1/2),(1/2)} = ((1/2)!)^{-2},$$

which does not make sense to anyone in the time of Wallis or Newton. So it looks as though his idea is doomed. It is at this point that Wallis shows his courage as a mathematician. He does not give up, and begins work on a new table, by optimistically adding new columns and rows for $p = \frac{1}{2}, \frac{3}{2}$ etc, and similarly for $q$. If only Wallis can numerically compute $a_{(1/2),(1/2)}$ (which is $\square$ in the next table), he will have achieved his goal of squaring the circle.

<table>
<thead>
<tr>
<th>$p\backslash q$</th>
<th>0</th>
<th>$\frac{1}{2}$</th>
<th>1</th>
<th>$\frac{3}{2}$</th>
<th>2</th>
<th>$\frac{5}{2}$</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\frac{1}{2}$</td>
<td>1</td>
<td>$\square$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>$\frac{3}{2}$</td>
<td>2</td>
<td>$\frac{5}{2}$</td>
<td>3</td>
<td>$\frac{7}{2}$</td>
<td>4</td>
</tr>
<tr>
<td>$\frac{3}{2}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$\frac{15}{8}$</td>
<td>3</td>
<td>$\frac{25}{8}$</td>
<td>6</td>
<td>$\frac{63}{8}$</td>
<td>10</td>
</tr>
<tr>
<td>$\frac{5}{2}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$\frac{35}{16}$</td>
<td>4</td>
<td>$\frac{105}{16}$</td>
<td>10</td>
<td>$\frac{231}{16}$</td>
<td>20</td>
</tr>
</tbody>
</table>

He can fill in the entries as long as $p$ is an integer, using formulas like (6) and (7), which make sense for any $q$. For example,

$$a_{2,1} = \frac{1}{2} \left( \frac{1}{2} + 1 \right) \left( \frac{1}{2} + 2 \right) = \frac{15}{8},$$
\[a_{3,\frac{3}{2}} = \frac{1}{6} \left( \frac{5}{2} + 1 \right) \left( \frac{5}{2} + 2 \right) \left( \frac{5}{2} + 3 \right) = \frac{231}{16}.\]

(Wallis is guessing here- he does not know for sure that formulas (6) and (7) actually give the value of the integral when \(p\) or \(q\) are not integers. Luckily, they do.) Then he notices the formula

\[a_{p,q} = \frac{p + q}{q} a_{p,q-1}.\]  

(8)

Here, we have only a recursive formula, but at least both sides make sense for any numbers \(p, q\). Again, Wallis only knows (8) for integers (you can check it yourself, using factorials) and he just assumes that (8) is true for non-integers. With these scruples happily tossed, formula (8) lets him move two steps to right in any row (remember that the steps are now by halves), and this allows him to fill in the blank spaces in table two, using \(\Box\). In row \(p = \frac{1}{2}\), for example, he gets

\[a_{\frac{1}{2},\frac{3}{2}} = \frac{1}{2} + \frac{3}{2} a_{\frac{1}{2},\frac{1}{2}} = \frac{4}{3} \Box.\]

Continuing like this, he eventually completes row \(p = \frac{1}{2}\) as follows

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
p & 0 & \frac{1}{2} & 1 & \frac{3}{2} & 2 & \frac{5}{2} & 3 \\
\hline
\frac{1}{2} & 1 & \Box & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & \frac{7}{2} \\
\hline
\end{array}
\]

Wallis remembers these entries are reciprocals of integrals, and the integrals get smaller as \(q\) increases, so the entries get larger as \(q\) increases. So for example, we have

\[\frac{2 \cdot 4}{5} \Box < \frac{3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6} < \frac{2 \cdot 2 \cdot 4 \cdot 4}{3 \cdot 5 \cdot 7} \Box,\]

Which, after remembering that \(\Box = \frac{4}{\pi}\), can be written

\[\frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6}{3 \cdot 3 \cdot 5 \cdot 5 \cdot 7} < \frac{\pi}{2} < \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdot \left( \frac{8}{7} \right)}{3 \cdot 3 \cdot 5 \cdot 5 \cdot 7} .\]

Thus we see Wallis’ formula emerging. The factor \(\frac{8}{7}\) says this approximation of \(\frac{\pi}{2}\) is accurate to within a factor of \(\frac{1}{7}\). If you go farther out in the row, the error becomes smaller than any given number. Wallis did not have the idea of limit, but instead invoked the argument of Euclid that two quantities whose difference is less than any given number are equal.

Incidentally, Wallis’ value for \(\Box\) seems to imply the mysterious-looking formula

\[(1/2)! = \frac{1}{\sqrt{\Box}} = \frac{\sqrt{\pi}}{2}.\]  

(9)

This will be explained later by Euler.

Though he gets credit for the discovery, is there not some question whether Wallis actually proved his formula? After all, he did make some reckless assumptions along the way. In fact, Wallis’ intuition and assumptions turn out to be correct. Even his method can be made rigorous, as we show in the next section.
2 The basic integrals

Let us revisit the $p = 1/2$ row in Wallis’ table. It’s $q^{th}$ term is

$$\int_0^1 (1 - x^2)^q \, dx.$$ 

we are especially interested in the case where $q = m/2$ is a half-integer. Making the substitution $x = \cos \theta$, we get

$$\int_0^1 (1 - x^2)^{m/2} \, dx = \int_0^{\pi/2} \sin^{m+1} x \, dx.$$

It is convenient to shift the indices by letting $n = m + 1$ and to consider the integrals

$$I_n = \int_0^1 (1 - x^2)^{(n-1)/2} \, dx = \int_0^{\pi/2} \sin^n x \, dx, \quad n = 0, 1, 2, 3, \ldots \tag{10}$$

The latter integrals are the areas under the graphs of $y = \sin^n x$, as shown below.

![Graph of \(y = \sin^n x\)](image)

Recall that $I_0 = \pi/2$ is the number we really want to compute in rational terms. We start by analyzing the entire family of integrals $I_n$. Even before computing these integrals $I_n$ we can make a qualitative observation, namely that

$$I_{n+1} < I_n. \tag{11}$$

This holds because $0 \leq \sin x \leq 1$ on the interval $0 \leq x \leq \frac{\pi}{2}$. It follows that for any positive integer $n$, we have

$$\sin^{n+1} x \leq \sin^n x. \tag{12}$$

There is therefore less area under the graph of $\sin^{n+1} x$ than under $\sin^n x$, so $I_{n+1} < I_n$, as claimed in (11).
On the other hand, we can almost compute the integrals \( I_n \). The results depend on whether \( n = 2k \) is even or \( n = 2k + 1 \) is odd.

It turns out (see next section for details) that

\[
I_{2k} = \left( \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdots \frac{2k-1}{2k} \right) \cdot \frac{\pi}{2} \quad \text{and} \quad I_{2k+1} = \left( \frac{1}{1} \cdot \frac{4}{3} \cdot \frac{6}{5} \cdots \frac{2k}{2k-1} \right) \cdot \frac{1}{2k+1}.
\]

The long terms in parentheses will occur repeatedly in these notes, so we will give them a name: for any positive integer \( k \) let us define \( P_k \) to be the product of the first \( k \) odd integers, divided by the product of the first \( k \) even integers:

\[
P_k = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdots \frac{2k-1}{2k}.
\]

We also define \( P_0 = 1 \).

With this new notation, our integrals can be expressed more concisely as

\[
I_{2k} = P_k \cdot \frac{\pi}{2} \quad I_{2k+1} = \frac{1}{P_k} \cdot \frac{1}{2k+1}.
\]

Note that Wallis’ fractions can also be expressed in terms of \( P_k \):

\[
W_k = \frac{1}{P_k^2} \cdot \frac{1}{2k+1}
\]

so all the parts are fitting together.

Now recall that Wallis used inequalities from three successive entries in his last table. We can do the same thing with our integrals: Applying (11) twice we have

\[
I_{2k+1} \leq I_{2k} \leq I_{2k-1}.
\]

Using the formulas (13) for \( I_{2k+1}, I_{2k}, I_{2k-1} \) respectively, this chain of inequalities says

\[
\frac{1}{P_k} \cdot \frac{1}{2k+1} \leq P_k \cdot \frac{\pi}{2} \leq \frac{1}{P_{k-1}} \cdot \frac{1}{2k-1}.
\]

Multiplying by the reciprocal of the leftmost term and using (14), we get

\[
1 \leq \frac{\pi}{2W_k} \leq \frac{P_k}{P_{k-1}} \cdot \frac{2k+1}{2k-1} = \frac{2k-1}{2k} \cdot \frac{2k+1}{2k-1} = 1 + \frac{1}{2k}.
\]

As \( k \to \infty \), we have \( 1/2k \to 0 \), and \( \pi/(2W_k) \) is trapped inside, so by the Squeeze Law, \( \pi/(2W_k) \to 1 \) as well. It follows that

\[
\lim_{k \to \infty} W_k = \frac{\pi}{2}.
\]

\[\text{We use the letter “P” because later we will see that } P_k \text{ is the probability of certain coin-tossing outcomes.} \]
as we wished to show. This completes the proof of Wallis’ formula.

To summarize, there were three main steps in the proof.

1. *An observation:* The number we want, $\pi/2$, is the integral $I_0$ which belongs to the family of integrals $I_n$.

2. *A qualitative inequality:* The integrals $I_n$ decrease as $n$ increases, because their integrands decrease.

3. *A computation:* We computed all the even integrals $I_{2k}$ in terms of $I_0$, and we computed the odd integrals $I_{2k+1}$ exactly.

This led to the unknown integral $I_0$ being squeezed between known integrals $I_{2k+1}$ and $I_{2k-1}$, from which we found a rational limit expression for $I_0$.

In doing all of this we have followed Wallis’ steps, but with more precision, because we had the formulas (13) for the integrals $I_n$.

### 2.1 Calculating the integrals $I_n$

In this section we recall how to calculate the integrals $I_n$. Recall that we have to show:

$$I_{2k} = \left(\frac{1 \cdot 3 \cdot 5 \cdots 2k-1}{2 \cdot 4 \cdot 6 \cdots 2k}\right) \cdot \frac{\pi}{2} \quad \text{and} \quad I_{2k+1} = \left(\frac{2 \cdot 4 \cdot 6 \cdots 2k}{1 \cdot 3 \cdot 5 \cdots 2k-1}\right) \cdot \frac{1}{2k+1} \quad (16)$$

At the beginning, it doesn’t matter if $n$ is even or odd. Starting from the definition

$$I_n = \int_0^{\pi/2} \sin^n x \, dx$$

we use integration by parts with

$$u = \sin^{n-1} x, \quad dv = \sin x \, dx, \quad du = (n-1) \sin^{n-2} x \cos x \, dx, \quad v = -\cos x.$$

Since

$$uv\bigg|_0^{\pi/2} = -\cos x \sin^{n-1} x \bigg|_0^{\pi/2} = 0,$$
the integration-by-parts formula gives
\[ I_n = \int_0^{\pi/2} \sin^n x \, dx = 0 - (n - 1) \int_0^{\pi/2} (\sin^{n-2} x \cos x)(-\cos x) \, dx \]
\[ = (n - 1) \int_0^{\pi/2} \sin^{n-2} x (\cos^2 x) \, dx \]
\[ = (n - 1) \int_0^{\pi/2} \sin^{n-2} x (1 - \sin^2 x) \, dx \]
\[ = (n - 1) \int_0^{\pi/2} (\sin^{n-2} x - \sin^n x) \, dx \]
\[ = (n - 1) \int_0^{\pi/2} \sin^{n-2} x \, dx - (n - 1) \int_0^{\pi/2} \sin^n x \, dx \]
\[ = (n - 1)I_{n-2} - (n - 1)I_n. \quad (17) \]

Adding \((n - 1)I_n\) to both sides we get
\[ nI_n = (n - 1)I_{n-2}. \]

This gives us the recursion formula
\[ I_n = \frac{n - 1}{n} I_{n-2}. \quad (18) \]

To get it started, we have the initial values
\[ I_0 = \int_0^{\pi/2} dx = \frac{\pi}{2}, \quad I_1 = \int_0^{\pi/2} \sin x \, dx = 1. \]

Then the recursion formula (18) takes over:
\[ I_2 = \frac{2 - 1}{2} I_0 = \frac{1}{2} \cdot \frac{\pi}{2}, \]
\[ I_4 = \frac{4 - 1}{4} I_2 = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{\pi}{2}, \]
\[ I_6 = \frac{6 - 1}{6} I_4 = \frac{5}{6} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{\pi}{2}, \]

and so on for \(I_{2k}\). And for the odds we have
\[ I_3 = \frac{3 - 1}{3} I_1 = \frac{2}{3} \cdot \frac{1}{3}, \]
\[ I_5 = \frac{5 - 1}{5} I_3 = \frac{4}{5} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{5}, \]
\[ I_7 = \frac{7 - 1}{7} I_5 = \frac{6}{7} \cdot \frac{4}{3} \cdot \frac{2}{1} \cdot \frac{1}{5} \cdot \frac{5}{7}, \]

and so on. This proves the formulas (16).

The following exercises are for practice only, not to hand in (except those on hw 7).
Exercise 1.1: Use \( u = \pi/2 - x \) and the identity \( \cos(x) = \sin(\pi/2 - x) \) to show that
\[
\int_0^{\pi/2} \cos^k x \, dx = \int_0^{\pi/2} \sin^k x \, dx.
\]
Then show that
\[
\int_0^\pi \sin^k x \, dx = 2 \int_0^{\pi/2} \sin^k x \, dx.
\]

Exercise 1.2: Using the substitution \( u = 1 - 2x \), show that
\[
\int_0^1 (x - x^2)^q \, dx = \frac{1}{4^q} \int_0^1 (1 - x^2)^q \, dx
\]
for any \( q \geq 0 \). Now suppose \( m \geq 0 \) is an integer, and calculate
\[
\int_0^1 (x - x^2)^{m/2}.
\]
Check your result for \( m = 1 \) by considering the graph of \((x - x^2)^{1/2}\).

Exercise 1.3: Here’s another approach to \( I_{2k} \), using complex numbers and Euler’s formula \( e^{ix} = \cos x + i \sin x \). Note that \( \sin x = (e^{ix} - e^{-ix})/2i \).

a) Compute \( \int_0^\pi e^{imx} \, dx \) for any integer \( m \).

b) Compute \( \int_0^\pi (e^{ix} - e^{-ix})^n \, dx \) by expanding the integrand.

c) Compute \( I_{2k} \). What happens for \( I_{2k+1} \)?

Exercise 1.4: Use the identity \( \cos^2 x = 1 - \sin^2 x \) to calculate
\[
\int_0^{\pi/2} \sin^6 x \cos^4 x \, dx
\]
(answer: \( I_6 - 2I_8 + I_{10} \). The same method works when the powers on \( \sin x \) and \( \cos x \) are both even. If, say, \( \cos x \) appears with odd power, you can split off a \( \cos x \), write the rest of the \( \cos x \)'s in terms of \( \sin^2 x \), and use the substitution \( u = \sin x \), instead of \( I_n \).)

Exercise 1.5: Letting \( u = 1 - x \), it is easy to show that
\[
\int_0^1 x^a(1 - x)^b \, dx = \int_0^1 x^b(1 - x)^a \, dx.
\]
If \( a \) or \( b \) is an integer, you can expand \((1 - x)\) in one of these integrals, and integrate term-by-term, for example:
\[
\int x^2(1-x)^{3/2} \, dx = \int_0^1 x^{3/2}(1-x)^2 \, dx = \int_0^1 x^{3/2} - 2x^{5/2} + x^{7/2} \, dx = \frac{2}{5} - \frac{16}{7} + \frac{2}{9}.
\]

But how to do it if neither \(a\) nor \(b\) are integers? Let \(x = \sin^2 \theta\) and show that
\[
\int_0^1 x^a(1-x)^b = \int_0^{\pi/2} \cos^{2a+1} \theta \sin^{2b+1} \theta \, d\theta.
\]

Then use the method of exercise 4 to calculate
\[
\int_0^1 x^{5/2}(1-x)^{3/2} \, dx.
\]

**Exercise 1.6:** Let \(p > 1\) be a constant. Show that
\[
\int_0^\infty \frac{1}{(1+x^2)^p} \, dx = \int_0^{\pi/2} \cos^{2p-2} \theta \, d\theta
\]
and then calculate this explicitly for \(p = m/2\), where \(m > 1\) is an integer. What happens if \(m = 1\)?

**Exercise 1.7:** Prove the formula used to make Wallis’ first table:
\[
\frac{p! \, q!}{(p+q)!} = \int_0^1 (1-x^{1/p})^q \, dx,
\]
where \(p\) and \(q\) are positive integers. Hint: Think of \(p\) as fixed, and let \(A_q\) and \(B_q\) be the left and right sides, respectively. First show that \(A_1 = B_1\). Then show that
\[
A_q = \frac{q}{p+q} A_{q-1}, \quad \text{and} \quad B_q = \frac{q}{p+q} B_{q-1}.
\]

The result for \(A_q\) follows from the definition. Use integration by parts for \(B_q\).

**Exercise 1.8:** Use (15) and the fact that \(\pi < 4\) to show that
\[
0 < \pi - W_n < \frac{2}{n+1}.
\]
Check this with \(W_{1000}\) as computed above.

**Exercise 1.9:** Here is a similar way to approximate \(e\) with fractions. Let
\[
L_n = \int_1^e (\ln x)^n \, dx.
\]
Use integration by parts to show that \(L_n = e - nL_{n-1}\). Then show that \(L_n \to 0\) by looking at the graph of \(\ln x\), and conclude that \(nL_{n-1} \to e\). Calculate a few \(L_n\)’s and approximate \(e\).

**Exercise 1.10:** What number is this?
\[
\left( \frac{2}{1} \right)^{1/2} \left( \frac{2 \cdot 4}{3 \cdot 3} \right)^{1/4} \left( \frac{4 \cdot 6 \cdot 8}{5 \cdot 5 \cdot 7} \right)^{1/8} \cdots
\]

12
3 Wallis and coin-tossing

The number $\pi$ is used for more than measuring circles, because it appears in many different areas of mathematics. Likewise, the Wallis formula for $\pi$ has many applications. In this section, we show how Wallis is related to sequences of random 0/1 events, like coin-tossing. This includes a probabilistic interpretation of the integrals $I_n$ used to prove Wallis’ formula.

First, we need a bit of background on binomial coefficients. Recall that the factorial of a positive integer $n$ is defined as

$$n! = n(n - 1)(n - 2) \cdots 2 \cdot 1.$$ 

We also define $0! = 1$. Why define $0!$ this way? For now, just accept that we define $0! = 1$ to make the formulas come out right. We will give a better reason in the next section.

There are many interpretations of $n!$. It is the number of ways to:

- put $n$ letters in $n$ mailboxes,
- arrange $n$ people in a row,
- paint $n$ houses with $n$ colors,
- permute $n$ distinct objects.

Now, if $n$ and $k$ are integers with $0 \leq k \leq n$, we define the binomial coefficient by

$$\binom{n}{k} = \frac{n!}{k!(n - k)!},$$

and call it “$n$ choose $k$”, because $\binom{n}{k}$ is the number of ways to choose $k$ objects from $n$ objects. From a group of $n$ people, you can form $\binom{n}{k}$ possible teams of $k$ members. For example, from a class of 30 people, you can make

$$\binom{30}{5} = \frac{30!}{5!25!} = 142506$$

possible basketball teams. Proof: make the class line up in all $30!$ possible ways, and each time take the first five for your team. You will get all possible teams this way, but you will get the same team several times, so we have to divide $30!$ by the number of times each team occurs. We get the same team from different lines by either permuting the first 5 members of the line or permuting the 25 members in the rest of the line. Thus, a total of $5!25!$ lines give the same team.
There are many other interpretations of binomial coefficients. For example, in algebra, we have the binomial expansion
\[(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k, \tag{19}\]
because the coefficient of \(x^k\) in \((1 + x)^n\) is the number of ways to choose \(k\) \(x\)'s from the \(n\) factors \((1 + x)\).

The Wallis formula has to do with the following interpretation of \(\binom{n}{k}\). If we label our \(n\) objects as \(1, 2, \ldots n\), then a choice of \(k\) objects can be expressed as a sequence of \(k\) 1's and \(n - k\) 0's, where the 1's correspond to the chosen objects. For example, the \(\binom{4}{2} = 6\) teams of two from a group of 4 are the sequences
\[1100, \ 1010, \ 1001, \ 0110, \ 0101, \ 0011.\]
Such sequences are also the outcomes of an experiment of \(n\) coin tosses, where 1 means heads and 0 means tails. For example, if we toss the coin 4 times, and get heads, tails, tails, heads, this outcome is 1001. Thus, when we toss the coin \(n\) times, the number of possible \(k\)-head outcomes is \(\binom{n}{k}\). The probability of getting \(k\) heads is
\[
\frac{\text{number of possible } k \text{ head outcomes}}{\text{number of all possible outcomes}} = \frac{1}{2^n} \binom{n}{k}.
\]
In particular, the probability of getting \(k\) heads from \(2k\) tosses is the number we will call \(P_k\), defined by
\[
P_k := \frac{1}{2^{2k}} \frac{(2k)!}{k!} = \frac{1}{2^{2k}} \frac{(2k)!}{(k!)^2} = \frac{(2k - 1)!!}{(2k)!!}.
\]
We have seen this number before. The formula for \(I_{2k}\) in (13) may be written as
\[
\frac{2}{\pi} \int_0^{\pi/2} \sin^{2k} x \, dx = P_k.
\]
This means that \(P_k\) is also the average of the function \(\sin^{2k} x\) on the interval \([0, \frac{\pi}{2}]\). Recall that in the proof of Wallis formula, we used the fact that \(\sin^{2k} x \to 0\) as \(k \to \infty\). Hence
\[
\lim_{k \to \infty} P_k = 0.
\]
This means that the probability of getting half heads in a large even number of tosses is essentially nil. This may seem strange, since half-heads is the most likely outcome. But there are more and more outcomes that take their share of the probability. We will examine this more closely in later sections.

The number \(P_k\) is given explicitly above. For large \(k\) it is very small, but its numerators and denominators are very big. This means we cannot compute \(P_k\) in practice. However, the Wallis formula may be viewed as an approximation to \(P_k\) for large \(k\). Recall that Wallis says that
\[
\lim_{k \to \infty} W_k = \frac{\pi}{2}.
\]
where
\[ W_k = \frac{2 2 4 4 6 6}{1 3 3 5 5 7} \cdots \frac{2k}{2k - 1} = \frac{2k}{2k + 1}. \]

Note that
\[ \frac{1}{W_k} = (2k + 1)^2. \]

So Wallis' formula for \( \pi \) can be written in the **probabilistic form**

\[
\lim_{k \to \infty} P_k \sqrt{2k + 1} = \sqrt{\frac{2}{\pi}}.
\]

This means that for large \( k \), we have the approximation

\[ P_k \approx \frac{c}{\sqrt{2k + 1}}, \]

where \( c = \sqrt{2/\pi} \) is a constant. Thus, Wallis tells us *how fast* the odds of getting the most likely outcome goes to zero.

**Exercise 2.1** Explain why it is a good idea to define \( 0! = 1 \), by giving a coin-tossing interpretation of \( \binom{n}{0} \) and \( \binom{n}{n} \).

**Exercise 2.2** Use the formula for \( \binom{n}{k} \) to prove that
\[
\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.
\]

**Exercise 2.3** Explain in two ways why
\[
\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n:
\]
first, using the formula for \( (1 + x)^n \), then using the coin tossing interpretation.

**Exercise 2.4** The **binomial triangle** (often called “Pascal’s triangle” even though it was known in various parts of the world many centuries before Pascal) has \( \binom{n}{k} \) in the \( k^{th} \) entry of row \( n \) from the top. Starting at the top of the binomial triangle, and moving downward at each step, what is the number of ways to get to the entry \( \binom{n}{k} \)? (Hint: interpret a choice of path as a sequence of coin-tosses.)

**Exercise 2.5** Cancelling \( (n - k)! \) from \( n! \), we can write
\[
\binom{n}{k} = \frac{1}{k!} n(n-1) \cdots (n-k+1),
\]
an expression which makes sense for any number $n$ and positive integer $k$. Show that
\[
\left(\frac{-1/2}{k}\right) = (-1)^k P_k, \quad \text{and} \quad \left(\frac{1/2}{k}\right) = (-1)^{k-1} \frac{P_k}{2k-1}.
\]

**Exercise 2.6** The Binomial Theorem, proved by Isaac Newton, is the expansion
\[
(1 + x)^q = \sum_{k=0}^{\infty} \binom{q}{k} x^k,
\]
which is valid for any number $q$ and $|x| < 1$. Show that formula (21) is consistent with formula (19) when $q$ is an integer $\geq 0$.

**Exercise 2.7** Show that
\[
\arcsin x = \sum_{k=0}^{\infty} P_k \frac{x^{2k+1}}{2k+1}.
\]
Hint: Use the fact that $\arcsin x = \int (1 - x^2)^{-1/2}$.

The next four exercises follow Euler, who used the series for $\arcsin x$ to calculate the sum $1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots$. There is a long story behind this sum, and Euler’s calculation of it made him famous. More about this sum later.

**Exercise 2.8** Show that
\[
\int_0^1 \frac{x^{2k+1}}{\sqrt{1 - x^2}} \, dx = \frac{1}{P_k(2k+1)}.
\]
Hint: make the substitution $u = x^2$ and use results from chapter 1.

**Exercise 2.9** Use the previous two exercises to show that
\[
\int_0^1 \frac{\arcsin x}{\sqrt{1 - x^2}} \, dx = 1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots.
\]

**Exercise 2.10** Show that
\[
\int_0^t \frac{\arcsin x}{\sqrt{1 - x^2}} \, dx = \frac{1}{2}(\arcsin t)^2.
\]

**Exercise 2.11** Combine the previous two exercises to compute the sum
\[
1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots.
\]
4 The Gaussian integral and factorial of one-half

We have been talking about factorials of integers, which are the building blocks of binomial coefficients. But we also saw that Wallis’ pursuit of his formula for $\pi$ led him to repeated confrontations with the strange number $(1/2)!$, leading to the equation

$$(1/2)! = \sqrt{\pi}/2$$

which is the version of Wallis’ formula that we mentioned in (9). So we have two questions:

1. What is $x!$ if $x$ is not an integer?

2. Assuming we understand $(1/2)!$, why is it equal to $\sqrt{\pi}/2$?

The answer to the first question was given by Euler (1707-1783). 4

Recall that for a positive integer $n$, the factorial $n!$ is defined as the product

$$n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 \quad (23)$$

of all positive integers $\leq n$. Formula (23) only makes sense if $n$ is a positive integer. We also define $0! = 1$, for some reason. The true definiton of $n!$ makes sense for any real number $n > -1$, not just integers.

$$n! = \int_0^\infty x^n e^{-x} \, dx. \quad (24)$$

For example,

$$0! = \int_0^\infty x^0 e^{-x} \, dx = -e^{-x}\bigg|_0^\infty = 1,$$

so we get $0! = 1$ from a general formula, instead of by Royal Decree, as before. And using integration by parts, you can check that $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ if $n$ is a positive integer.

However, most interesting to us now is that formula (24) says

$$(1/2)! = \int_0^\infty x^{1/2} e^{-x} \, dx, \quad \text{and} \quad (-1/2)! = \int_0^\infty x^{-1/2} e^{-x} \, dx.$$

What are the values of these integrals? They are actually famous integrals in disguise, used in many areas of mathematics. Let’s work on the second one starting with the substitution $u = x^{1/2}$. We get $dx = 2u \, du$, so

$$(-1/2)! = \int_0^\infty x^{-1/2} e^{-x} \, dx = 2 \int_0^\infty u^{-1} e^{-u^2} \, u \, du = 2 \int_{-\infty}^\infty e^{-u^2} \, du.$$ 

4Euler also answered the second question using multivariable calculus, which is beyond the scope of this class. Fortunately we can use Wallis’ formula to answer question 2 without multivariable calculus.
so whatever it is, the number $(-1/2)!$ is the area under the whole graph of $e^{-x^2}$, which is the famous "Bell Curve" used in probability. And

$$\frac{1}{2}! = \frac{1}{2} \cdot (-1/2)! = \int_0^\infty e^{-x^2} \, dx$$

is half of this area.

We keep talking about $\frac{1}{2}!$ but have not actually computed it. In fact the integral (25) cannot be computed by finding an antiderivative of $e^{-x^2}$ (go on, try it!). If only there were an extra $x$, we could do it. Namely, if instead of the integral in (25), we had

$$\int_0^\infty xe^{-x^2} \, dx,$$

then taking $u = x^2$ would turn it into

$$\int_0^\infty xe^{-x^2} \, dx = \frac{1}{2} \int_0^\infty e^{-u} \, du = (1/2).$$

If we had an extra $x^2$, and did $u = x^2$ again, we’d get

$$\int_0^\infty x^2e^{-x^2} \, dx = \frac{1}{2} \int_0^\infty u^{1/2}e^{-u} \, du = (1/2) \cdot (1/2)!$$

which is back to the hard integral we started with. A more clever idea is to use integration by parts, with

$$u = x, \quad dv = xe^{-x^2},$$

since we just integrated $dv$. However, this will lead to the same hard integral again. But at least we get back to the same hard integral, and not some new one.

This reminds us of the story with Wallis' integrals $I_n$, which were easy for odd $n$ and hard for even $n$.

To pursue this path, let us define

$$G_n = \int_0^\infty x^n e^{-x^2} \, dx.$$
The integral \( G_0 = \int_0^\infty e^{-x^2} \, dx \) is the famous **Gaussian integral**, which is what we want to know.

So far, we only know that

\[
G_0 = (1/2)! =?, \quad G_1 = (1/2), \quad G_2 = (1/2) \cdot (1/2)! =? .
\]

Even though we know it won’t work completely, let’s try integration by parts on \( G_n \) for \( n \geq 2 \) and see what happens. With \( u = x^{n-1} \) and \( dv = xe^{-x^2} \, dx \), we get

\[
G_n = \int_0^\infty x^n e^{-x^2} \, dx
= \frac{1}{2} x^{n-1} e^{-x^2}\big|_0^\infty + \frac{n-1}{2} \int_0^\infty x^{n-2} e^{-x^2} \, dx
= \frac{n-1}{2} G_{n-2},
\]

since \( n \geq 2 \) and \( \lim_{x \to \infty} x^{n-1} e^{-x^2} = 0 \). So we have the recursion formula

\[
G_n = \frac{n-1}{2} G_{n-2}, \quad (26)
\]

with the initial values

\[
G_0 = (1/2)! \quad \text{(hard)} \quad G_1 = (1/2) \quad \text{(easy)}.
\]

This means the even \( G \)'s are hard, and the odd \( G \)'s are easy:

\[
G_{2k} = \frac{2k-1}{2} \cdot G_{2k-2}
= \frac{2k-1}{2} \cdot \frac{2k-3}{2} \cdot G_{2k-4}
= \cdots
= \frac{2k-1}{2} \cdot \frac{2k-3}{2} \cdots \frac{3}{2} \cdot \frac{1}{2} \cdot G_0
= k! \cdot P_k \cdot G_0,
\]

where we recall from the previous section that

\[
P_k = \frac{1}{4^k} \binom{2k}{k} = \frac{1 \cdot 3 \cdots (2k-1)}{2 \cdot 4 \cdots (2k)}
\]

is the probability of getting \( k \) heads in \( 2k \) coin-tosses. So all the even \( G \)'s boil down to the single hard integral \( G_0 \), just as the even \( I \)'s boiled down to \( I_0 \).
On the other hand, for the odd $G$'s, we get an actual answer:

$$G_{2k+1} = \frac{2k}{2} \cdot G_{2k-1}$$
$$= \frac{2k}{2} \cdot \frac{2k-2}{2} \cdot \frac{2k-4}{2} \cdot G_{2k-3}$$
$$\vdots$$
$$= \frac{2k}{2} \cdot \frac{2k-2}{2} \cdots \frac{2}{2} \cdot G_1$$
$$= \frac{k!}{2}, \text{ since } G_1 = 1.$$  \hfill (28)

Continuing the pattern of how we studied the Wallis integrals $I_n$, the next step is some kind of inequality, analogous to $I_{n+1} < I_n$. This is less obvious; it took another century for the correct inequality for $G_n$ to be discovered by Stieltjes.  \footnote{Thomas Stieltjes (1856-1894) was a Dutch mathematician who is today remembered mainly for the “Stieltjes integral”, which you would encounter in an advanced analysis class.} What follows uses only math that we already know, but is not very well-known. I found it by accident while looking for something else in Stieltjes’ *Complete Works*.

The new inequality is

$$G_n^2 < G_{n-1}G_{n+1}, \quad \text{for all } n \geq 1.$$  \hfill (29)

This is less obvious than the inequality $I_{n+1} < I_n$. I don’t know how Stieltjes thought of it, but once thought of, it is not hard to prove in the same way, as follows.

Where have we seen something like $B^2 - AC$? This is what goes inside the square root in the quadratic formula for the polynomial $p(t) = At^2 + 2Bt + C$, whose roots are $(-B \pm \sqrt{B^2 - AC})/A$. We will have $B^2 < AC$ precisely when the polynomial has no real roots, which happens precisely when the function $p(t) = At^2 + 2Bt + C$ never changes sign.

In our case, the quadratic polynomial that gives rise to the terms in (29) is

$$p(t) = t^2 G_{n-1} + 2tG_n + G_{n+1}.$$  

Our goal is to show that $p(t)$ never changes sign. Remember that the $G$’s are just numbers (which we happen not to know completely), and they are the coefficients in the polynomial $p(t)$. Of course, the $G$’s are also the values of integrals:

$$G_{n-1} = \int_0^\infty x^{n-1}e^{-x^2}dx, \quad G_n = \int_0^\infty x^n e^{-x^2}dx, \quad G_{n+1} = \int_0^\infty x^{n+1} e^{-x^2}dx,$$

so $p(t)$ is a sum of integrals. Let’s combine these into one integral (note that $t$ is a constant with
respect to $dx$, so can be moved inside the integrals):

\[
p(t) = t^2 \int_0^\infty x^{n-1}e^{-x^2} \, dx + 2t \int_0^\infty x^n e^{-x^2} \, dx + \int_0^\infty x^{n+1}e^{-x^2} \, dx
\]

\[
= \int_0^\infty (t^2 x^{n-1} + 2tx^n + x^{n+1})e^{-x^2} \, dx
\]

\[
= \int_0^\infty x^{n-1}(t + x)^2e^{-x^2} \, dx.
\]

This integral is the worst one yet, but we do not have to actually compute it. Just note that the integrand (as a function of $x$) is $\geq 0$ and is equal to zero at no more than two points. So there is positive area under the integrand, and the integral is positive. Hence,

\[
p(t) > 0 \quad \text{for all } \ t.
\]

As discussed above, this means $p(t)$ never changes sign, so $B^2 - AC < 0$. Thus we have proved that

\[
G_n^2 - G_{n-1}G_{n+1} < 0,
\]

which is the inequality (29) that we wanted to prove.

We now proceed just as with the integrals $I_n$. We insert our calculations for $G_n$ into the inequality (29) so as to get our desired integral $G_0$ squeezed between things we know. Then Wallis himself will compute the limit for us!

We insert the recursion

\[
G_n = \frac{n-1}{2}G_{n-2}
\]

into the inequality

\[
G_n^2 < G_{n-1}G_{n+1}.
\]

(31)

For both $n = 2k + 1$ and then $n = 2k$. This gives

\[
G_{2k+1}^2 < \frac{2k+1}{2}G_{2k}^2 < \frac{2k+1}{2}G_{2k-1}G_{2k+1}.
\]

(32)

Now plug our computations (27) and (28)

\[
G_{2k} = k!p_kG_0, \quad G_{2k+1} = \frac{k!}{2}
\]

into (32) to get

\[
\left(\frac{k!}{2}\right)^2 < \frac{2k+1}{2}(k!p_kG_0)^2 < \frac{2k+1}{2} \cdot \frac{(k-1)!}{2} \cdot \frac{k!}{2}.
\]

Dividing everything by $(k!/2)^2$, we get

\[
1 < 2(2k+1)p_k^2G_0^2 < \frac{2k+1}{2k} = 1 + \frac{1}{2k}
\]

(33)
We have seen that Wallis’ formula can be written
\[
\lim_{k \to \infty} (2k + 1)P_k^2 = \frac{2}{\pi},
\]
So taking \(k \to \infty\) in (33), we get
\[
\frac{4}{\pi} G_0^2 = 1,
\]
meaning that
\[
G_0 = \frac{\sqrt{\pi}}{2},
\]
and we are done.

Let’s review. Recall that \(G_0\) is the Gaussian integral, and is also \((1/2)!!\):
\[
G_0 = \int_0^\infty e^{-x^2} \, dx = (1/2)!
\]
By computing that \(G_0 = \sqrt{\pi}/2\), we have shown that
\[
\int_0^\infty e^{-x^2} \, dx = (1/2)! = \frac{\sqrt{\pi}}{2}.
\]
In other words, the area under the whole graph of the Bell Curve \(e^{-x^2}\) is exactly \(\sqrt{\pi}\). This computation combined the work of Wallis, Euler and Stieltjes, from the 17\(^{th}\), 18\(^{th}\) and 19\(^{th}\) centuries, respectively. The limits 0 and \(\infty\) on the integral were essential: We never computed the antiderivative \(\int e^{-x^2} \, dx\).

Since \(\pi\) is involved, you might guess that \(G_0^2\) is somehow related to the area of a circle. This is true, and leads to another way to compute \(G_0\), which is easier than what we just did (and was known long before Stieltjes), but requires double integrals and is beyond our course. In the other direction, knowing \(G_0\) allows one to compute the volume of a sphere in any dimension, using just single integrals (see section ??).

Since all the even \(G\)'s boiled down to \(G_0\), we have actually computed many other integrals. We leave this to the exercises.

**Exercise 3.1:** Use the recursion formula \(n! = n \cdot (n-1)!\) to compute \((-1/2)!\), \((3/2)!\), and \((5/2)!\) in terms of \(\sqrt{\pi}\).

**Exercise 3.2:** Give a formula for \((k - (1/2))!\) for any integer \(k \geq 0\).

**Exercise 3.3:** We have seen that \(G_0 = (1/2) \cdot (-1/2)!\). What is the analogous formula for \(G_{2k}\)? (Hint: use equation (30).)
In the remaining exercises, make a substitution to turn the integral into the factorial integral, then compute it. Leave all answers as fractions.

**Exercise 3.4:** Calculate $\int_{0}^{\infty} x^6 e^{-2x} \, dx$.

**Exercise 3.5:** Calculate $\int_{0}^{\infty} x^6 e^{-4x^2} \, dx$.

**Exercise 3.6:** Calculate $\int_{0}^{\infty} \sqrt{x} e^{-x^3} \, dx$.

**Exercise 3.7:** Calculate $\int_{0}^{\infty} 3^{-4x^2} \, dx$.

**Exercise 3.8:** Calculate $\int_{0}^{\infty} e^{-ax^2} \, dx$ ($a$ is a positive constant).

**Exercise 3.10:** Show that $\int_{0}^{\infty} e^{-xp} \, dx = \left( \frac{1}{p} \right)!$.

**Exercise 3.11:** Calculate $\int_{0}^{1} \frac{dx}{\sqrt{-\log x}}$.

## 5 Volumes of spheres

We began with the unit circle $C$, whose area is given by the integral

$$\text{Area}(C) = 4 \int_{0}^{1} \sqrt{1-x^2} \, dx = 4 \int_{0}^{\pi/2} \sin^2 \theta \, d\theta = 4I_2.$$  

Wallis’ essential observation was that $I_2$ belongs to the family of integrals

$$\int_{0}^{1} (1-x^2)^{(n-1)/2} \, dx = \int_{0}^{\pi/2} \sin^n \theta \, d\theta = I_n,$$

for $n = 0, 1, 2, 3, \ldots$. By studying this family, he discovered his product formula for $I_2$, thus obtaining a formula for $\pi$, as we saw in section (000).

Now the circle lives in the plane, which has dimension $n = 2$. What about $n = 3$? Archimedes found that the volume of the unit sphere in three dimensions is $4\pi/3$. Using Cavalieri’s principle, we can derive Archimedes’ volume formula by slicing the sphere with planes parallel to the axis through the north and south poles. Let the sphere have equation $x^2 + y^2 + z^2 = 1$, where the axis is the $z$-axis. For each fixed $z$, the slice is the circle with equation

$$x^2 + y^2 = 1 - z^2,$$
radius \sqrt{1-z^2}, and area \pi \cdot (\sqrt{1-z^2})^2 = \pi \cdot (1-z^2). Since the slices go from \( z = -1 \) (south pole) to \( z = 1 \) (north pole) the volume of the sphere \( S \) is

\[
\text{Vol}(S) = \int_{-1}^{1} \pi \cdot (1-z^2) \, dz = 2\pi \int_{0}^{1} (1-z^2) \, dz = 2\pi \cdot I_3.
\]

And we know that \( I_3 = 2/3 \), so this gives \( \text{Vol}(S) = 4\pi/3 \), as found by Archimedes.

Thus, for \( n = 2 \) the area is \( 4I_2 \), while for \( n = 3 \) the volume is \( 2\pi I_3 \). Note that \( 2\pi \) is twice the area of \( C \), and 4 is twice the length of the line segment \( L = [-1,1] \), which is the one-dimensional analogue of \( C \) and \( S \). For \( L, C \) and \( S \) consist of all points in 1, 2, 3 dimensions of distance at most one from the origin.

If we have an arbitrary radius \( r > 0 \), then \( L, C, S \) become the line segment of length \( 2r \), the inside of a plane circle of radius \( r \), and the inside of a sphere of radius \( r \); all three consist of the points in 1, 2 or 3-dimensional space whose distance from a fixed point (the center) is at most \( r \).

Each dimension has its own way of measuring size, namely length, area, volume. We know:

- the length of the line segment is \( 2r \)
- the area of the circle is \( \pi r^2 \)
- the volume of the sphere is \( (4\pi/3)r^3 \).

There is no reason to stop at three dimensions. The **hypersphere** of radius \( r \) is the collection of all points in 4-dimensional space having distance at most \( r \) from the origin. The “surface” of the hypersphere consists of the points \((x, y, z, w)\) satisfying

\[
x^2 + y^2 + z^2 + w^2 = r^2.
\]

The four-dimensional analogue of _length, area, volume, . . . _ is called **hypervolume**. We will add to our list above by showing that

the hypervolume of the hypersphere is \( (\pi^2/2)r^4 \).

Just as two-dimensional creatures cannot see a whole sphere in three dimensions, we three-dimensional creatures cannot see a whole hypersphere. However, we can see slices of the hypersphere. If we slice the hypersphere at a fixed value of \( w \), we see a 3-dimensional sphere with equation

\[
x^2 + y^2 + z^2 = r^2 - w^2
\]

and having radius \( \sqrt{r^2 - w^2} \). As \( w \) varies from \( -r \) to \( r \) we see a movie of slices. When \( w = -r \) the slice is just a point (the “south pole” of the hypersphere). When \( w = 0 \) we get the biggest slice: a sphere of radius \( r \) (the “equator”) and when \( w = r \) we just have a point again (the “north pole”).

24
We can again use Cavalieri’s principle to compute the hypervolume $V_4$ of the hypersphere of radius $r = 1$ by integrating the volumes of the slices. Since the slice at $w$ is a sphere of radius $\sqrt{1 - w^2}$, its volume is $(4/3)\pi \cdot (1 - w^2)^{3/2}$. Hence we get

$$V_4 = \int_{-1}^{1} \frac{4\pi}{3} \cdot (1 - w^2)^{3/2} \, dw = \frac{8\pi}{3} \int_{0}^{1} (1 - w^2)^{3/2} \, dw = \frac{8\pi}{3} I_4 =$$

Now, the factor $8\pi/3$ is twice the volume of the unit sphere in three dimensions. Since

$$I_4 = \left( \frac{1}{2} \cdot \frac{3}{4} \right) \cdot \frac{\pi}{2},$$

we find that the hypervolume of the unit hypersphere is

$$V_4 = \frac{\pi^2}{2}.$$

If we had an arbitrary radius $r$, the same computation would tell us that the volume is $(\pi^2/2)r^4$, as claimed above.

Let’s keep going. The a sphere of radius $r$ in $n$ dimensions has equation

$$x_1^2 + x_2^2 + \cdots + x_n^2 = r^2.$$

If we slice at a particular value of $x_n$, we get a sphere in $n - 1$-dimensions, with equation

$$x_1^2 + x_2^2 + \cdots + x_{n-1}^2 = r^2 - x_n^2,$$

which has radius $\sqrt{r^2 - x_n^2}$. So the $n - 1$-dimensional volume of the slice is

$$v_{n-1} (r^2 - x_n^2) = v_{n-1} (1) \cdot (r^2 - x_n^2)^{(n-1)/2}.$$

By Cavalieri’s principle, the volume of the $n$-dimensional sphere is the integral of the volumes of the slices. Let $V_n$ be the volume of the sphere of radius $r = 1$ in $n$ dimensions.

The same computation as above shows that

$$V_n = 2V_{n-1} \cdot I_n.$$

This is a recursion formula for $V_n$. The value of $I_n$ depends on whether $n$ is even or odd, in which cases $n - 1$ is odd or even. To avoid this back-and-forth, we apply the recursion formula one more time to $V_{n-1}$ and get

$$V_n = 4V_{n-2} \cdot I_n \cdot I_{n-1}.$$

Remarkably, $I_n \cdot I_{n-1}$ does not depend on the parity of $n$. Indeed, from our formulas for the integrals $I_n$ (see Wallis-Gauss notes) we have

$$I_n \cdot I_{n-1} = \frac{\pi}{2n} \quad \text{for all } n \geq 1.$$
So we get
\[ V_n = 4 \cdot \frac{\pi}{2^n} \cdot V_{n-2} = \frac{2\pi}{n} \cdot V_{n-2}. \]
This looks like the recursion for factorials. in fact, \( \pi^{n/2}/(n/2)! \) has the same recursion:
\[ \frac{\pi^{n/2}}{(n/2)!} = \frac{2\pi}{n} \cdot \frac{\pi^{(n-2)/2}}{((n-2)/2)!}. \]
And you can check that
\[
\begin{align*}
\frac{\pi^{1/2}}{(1/2)!} &= 2 = V_1 \\
\frac{\pi^{2/2}}{(2/2)!} &= \pi = V_2 \\
\frac{\pi^{3/2}}{(3/2)!} &= 4\pi/3 = V_3 \\
\frac{\pi^{4/2}}{(4/2)!} &= \pi^2/2 = V_4
\end{align*}
\]
Since both sequences \( V_n \) and \( \pi^{n/2}/(n/2)! \) have the same initial terms and the same recursion, it follows that
\[ V_n = \frac{\pi^{n/2}}{(n/2)!} \quad \text{for all } n \geq 1. \]
Taking an arbitrary radius \( r > 1 \) just gives a factor of \( r^n \). Thus, starting with the circle, Wallis’ integrals \( I_n \) have led to a uniform formula for spheres in any dimension:
the volume of a sphere of radius \( r \) in \( n \) dimensions is \( \frac{\pi^{n/2}}{(n/2)!} r^n \).

6 Wallis and the binomial distribution

If we graph the probabilities
\[ \frac{1}{2^n} \binom{n}{k} \]
of getting \( k \) heads in \( n \) coin tosses, we get pictures like
for $n = 4$ and $n = 6$. These are discrete versions of bell curves. In the previous section, we saw that the highest dot, which is the probability of the most likely outcome happening, goes down to zero as $n$ increases. Moreover, Wallis told us precisely how fast this dot goes to zero. In this section we will see that Wallis also tells us the parameters of the bell curve which approximates the binomial distributions above.

As we have mentioned, the basic bell curve is the graph of $e^{-x^2}$. This has a maximum of 1 at $x = 0$, and inflection points at $\pm 1/\sqrt{2}$.

The latter are the points where the graph of $e^{-x^2}$ begins to flare outwards. The point 0 is the mean, and $1/\sqrt{2}$ is the standard deviation of this bell curve.

However, we need bell curves with an arbitrary mean $\mu$ and standard deviation $\sigma$. Such a curve is the graph of

$$y = \exp\left(-\frac{1}{2} \left[ \frac{x - \mu}{\sigma} \right]^2 \right).$$

This is obtained from the graph of $e^{-x^2}$, but shifted to have its maximum at $\mu$, and stretched to have its inflection points at $\mu \pm \sigma$.

For example here is the graph with $\mu = \sigma = 1$. 

![Diagram of bell curve with mean and standard deviation](image)
Exercise 4.1: Show that the function in (34) has its maximum at $\mu$ and inflection points at $\mu \pm \sigma$.

Exercise 4.2: Use our formula for the Gaussian integral $\int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi}$ to show that

$$\int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} \left[ \frac{x - \mu}{\sigma} \right]^2 \right) \, dx = \sigma \sqrt{2\pi}.$$

The function

$$\frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{x - \mu}{\sigma} \right]^2 \right)$$

is called the Gaussian (or “normal”) distribution. It approximates, in an easily understood visual way, many different occurrences of discrete random behavior that may be very difficult to compute one at a time. You just have to adjust $\mu$ and $\sigma$ to the case at hand.

For example, it is very difficult to compute the binomial coefficients necessary to determine the exact proportion of outcomes with $k$ heads from $n$ coin tosses, if $n$ is large. Instead, we can use the approximation

$$\frac{1}{2^n} \binom{n}{k} \sim \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{k - \mu}{\sigma} \right]^2 \right) \quad (n \text{ large}). \quad (35)$$

We just have to determine $\mu$ and $\sigma$. The maximum of the right side of (35), which is $\mu$, should be the most likely outcome of the left side, which is $n/2$, so

$$\mu = \frac{n}{2}.$$

What about the standard deviation $\sigma$? By now it is clear that the Wallis formula knows everything about large binomial coefficients, so it is no surprise that Wallis will tell us $\sigma$. Taking $k = \mu = n/2$ in (35), we get the approximation

$$\frac{1}{2^n} \binom{n}{\mu} \sim \frac{1}{\sigma \sqrt{2\pi}}$$

for large $n$. On the other hand, by Wallis (see equation (20)), we have

$$\frac{1}{2^n} \binom{n}{\mu} \sim \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{n+1}} \sim \sqrt{\frac{2}{n\pi}}$$

for large $n$. So we should have

$$\frac{1}{\sigma \sqrt{2\pi}} = \sqrt{\frac{2}{n\pi}},$$

meaning that $\sigma = \sqrt{n}/2$. 

28
For example, if \( n = 6 \) we get \( \mu = 3 \) and \( \sigma = \sqrt{6}/2 \), so the approximating bell curve is

\[
\frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{k - \mu}{\sigma} \right]^2 \right) = \frac{1}{3\pi} \exp \left( -(x - 3)^2/3 \right).
\]

If we overlay this on the plotpoints of \( \frac{1}{2^n} \binom{n}{k} \) it already looks pretty good, even for this small \( n \).

\[
y = \frac{1}{\sqrt{3\pi}} e^{-(x-3)^2/3}
\]

In summary, for large \( n \), we have the approximation (as functions of \( k \))

\[
\frac{1}{2^n} \binom{n}{k} \sim \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{k - \mu}{\sigma} \right]^2 \right) \quad \text{where} \quad \mu = \frac{n}{2}, \quad \sigma = \frac{\sqrt{n}}{2}.
\]

This is one of many instances of a discrete function being approximated by a continuous function. It may seem paradoxical, but the latter is easier to work with, as we will see in the next chapter.

7. The function \( \text{erf}(x) \).

We have seen that with a large number \( n \) of coin tosses the probability of getting any particular outcome is almost nil. One is more interested in the probability that a certain range of outcomes will occur. For example, if we toss a coin 100 times, the probability of getting exactly 50 heads is almost zero, but what is the probability of getting between 48 and 52 heads? To answer this we could compute 5 very large binomial coefficients, add them up, and divide by \( 2^{100} \). It is much easier to use our formula (36) in the previous section.

The probability of getting between \( a \) and \( b \) heads in \( n \) coin tosses is exactly

\[
\frac{1}{2^n} \sum_{k=a}^{b} \binom{n}{k}.
\]

By (36), this probability is approximately

\[
\frac{1}{\sigma \sqrt{2\pi}} \int_a^b \exp \left( -\frac{1}{2} \left[ \frac{x - \mu}{\sigma} \right]^2 \right) \, dx
\]

(37)
To handle this integral, we define the **error function**:

\[
\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} \, dt.
\]

Here \( x \) can be any real number. So \( \text{erf}(x) \) is the part of the area under a certain bell curve, between 0 and \( x \).

**Exercise 5.1:** Show that \( \text{erf}(x) \) has the following properties.

1. \( \text{erf}(0) = 0 \).
2. \( \lim_{x \to \infty} \text{erf}(x) = \frac{1}{2} \).
3. \( \text{erf}(x) \) is an odd function. That is, \( \text{erf}(-x) = -\text{erf}(x) \).
4. \( \text{erf}(x) \) is always increasing, is concave up for \( x < 0 \), and concave down for \( x > 0 \).

**Exercise 5.2:** Show that

\[
\frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp \left( -\frac{1}{2} \left[ \frac{k-\mu}{\sigma} \right]^2 \right) \, dx = \text{erf} \left( \frac{b-\mu}{\sigma} \right) - \text{erf} \left( \frac{a-\mu}{\sigma} \right).
\]

Thus, if you make a large number \( n \) of tosses, the probability of getting between \( a \) and \( b \) heads is approximately

\[
\text{erf} \left( \frac{b-\mu}{\sigma} \right) - \text{erf} \left( \frac{a-\mu}{\sigma} \right), \quad \mu = \frac{n}{2}, \quad \sigma = \frac{\sqrt{n}}{2}.
\]

Here is a table of values of \( \text{erf}(x) \) for small \( x \) (for larger \( x \), we have \( \text{erf}(x) \simeq .5 \)).

<table>
<thead>
<tr>
<th>( x )</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{erf}(x) )</td>
<td>.0398</td>
<td>.0793</td>
<td>.1179</td>
<td>.1554</td>
<td>.1915</td>
<td>.2257</td>
<td>.2580</td>
<td>.2881</td>
<td>.3159</td>
<td>.3413</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( x )</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{erf}(x) )</td>
<td>.3643</td>
<td>.3849</td>
<td>.4032</td>
<td>.4192</td>
<td>.4332</td>
<td>.4452</td>
<td>.4554</td>
<td>.4641</td>
<td>.4713</td>
<td>.4773</td>
</tr>
</tbody>
</table>

**Example:** If we toss 100 times, what is the approximate probability of getting between 48 and 52 heads? We have

\[
n = 100, \quad \mu = 50, \quad \sigma = 5,
\]

---

\(^6\)This is how Gauss defined \( \text{erf}(x) \). Most other authors define \( \text{Erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} \, dt \). The two versions are related by \( 2\text{erf}(x) = \text{Erf}(x/\sqrt{2}) \). Our \( \text{erf}(x) \) is slightly more complicated, but produces simpler formulas in the end.
so the approximate probability is
\[
\text{erf} \left( \frac{52 - 50}{5} \right) - \text{erf} \left( \frac{48 - 50}{5} \right) = 2 \text{erf}(.4) \simeq .3108 \ldots
\]

With these 100 tosses, what is the approximate probability of getting at least 40 heads? Here, 
\(n, \mu, \sigma\) are unchanged, but \(a = 40, \ b = \infty\), so the approximate probability is
\[
\text{erf} \left( \frac{\infty - 50}{5} \right) - \text{erf} \left( \frac{40 - 50}{5} \right) = \text{erf}(\infty) - \text{erf}(-2) \sim .5 + .4773 \sim .9773 \ldots
\]

This example is small enough that we can compute the actual probabilities. We get
\[
\sum_{k=48}^{52} \frac{1}{2^{100}} \binom{100}{k} \sim .3827 \ldots
\]
\[
\sum_{k=40}^{100} \frac{1}{2^{100}} \binom{100}{k} \sim .9824 \ldots
\]

Our approximation .3108 for the first one is quite poor, even though the Gaussian distribution is an excellent approximation to the actual distribution. To see the problem, think of the sum
\[
\sum_{k=48}^{52} \frac{1}{2^{100}} \binom{100}{k}
\]
as a Riemann sum of five rectangles of width 1 on the intervals \([48, 49], [49, 50], \ldots\). The last rectangle sits on \([52, 53]\), which is not accounted for in the integral; this last rectangle causes the large error. If we integrate from 48 to 53 we get \(\text{erf}(.6) + \text{erf}(.4) = .3811\), which is much better. The second example is closer because there are no extra rectangles.

We will not worry about this any further in this course.

**Exercise 5.3:** For \(n = 100\) tosses, find the approximate probabilities of getting
a) between 45 and 55 heads (answer: .6826)
b) between 50 and 60 heads
c) at least 45 heads

**Exercise 5.4:** For \(n = 10,000\) tosses, find the approximate probabilities of getting
a) between 4950 and 5050 heads (answer: .6826 again)
b) between 4900 and 5100 heads
c) no more than 4500 heads
Exercise 5.5: In this problem, you are not given $n$, so you won’t know $\mu$ and $\sigma$ either. Nevertheless, please find the approximate probabilities of getting

a) between $\mu - \sigma$ and $\mu + \sigma$ heads

b) between $\mu - 2\sigma$ and $\mu + 2\sigma$ heads.

Besides coin tossing, the same method can compute approximate probabilities in any situation where there are two possible results, equally likely, and the trial is repeated a large number of times.
Exercise 5.6: (A story problem) The town of Bumpkin has the shape of a triangle, with BankBumpkin at the northern peak, in Upper Bumpkin, where all the money resides. The streets of Bumpkin were made long ago by well-organized cows, and the map of Bumpkin looks like this:

```
BB
 /\
/ / /\ 
/ / / / \
/ / / / / \
:\:\:\:\:\:\:\:
```

Below Upper Bumpkin lies Middle Bumpkin, and even further south is Lower Bumpkin (not shown), where the map is very complicated indeed.

Jack and Jill grew up in Lower Bumpkin. After Jack fell down and broke his crown, he was never the same, and, after a series of Failures in Life, poor Jack had turned to a life of crime. And so one day, Jack went north to rob BankBumpkin. He tied up everyone in the bank, took all the cash he could find, and dashed out the door, making a run for it down the streets toward the labyrinth of Lower Bumpkin. Jack knew that if he got far enough south, the cops would never find him. However, one of the tellers recognized Jack, broke loose and called the police.

As for Jill, who came tumbling down after Jack in that famous accident, she had made a full recovery, and being very clever, went on to become a Pure Mathematician. But since Jill was interested neither in teaching nor in practical applications of mathematics, she was forced to support herself as a police dispatcher, which in the usually quiet town of Bumpkin allowed plenty of time for research, and many of her mathematical discoveries were made while contemplating the map of Bumpkin on the station wall. And it was Jill who answered the phone after the bank robbery.

Jill knew Jack pretty well from the old days. She knew he could run fast, and that he was thinking only of getting to Lower Bumpkin as quickly as possible. Also he was surely panicked, making a random choice at each intersection, though always heading in a southerly direction. So Jill made some brief calculations on police stationary. She figured that by the time the police got rolling, Jack would be nearing his 100th intersection. But which one? There were 101 possibilities, and not nearly that many cops on the Bumpkin beat, so Jill suggested a deployment of just enough police to have a 95 percent chance of catching Jack. With one uniform at each intersection, how many police did she deploy?
If $ab$ is an hyperbola, 
$$x = c, \quad y = \frac{1}{x}.$$  
If also, 
$$C = 1, \quad a = 1 + x, \quad b = 1 + \frac{1}{x}.$$  
Then, shall the area of these

As in this table, so will $y$ first area be also inserted. The composition of this table may be

Suppose $a, b, c$ is a square

The sum of any figure is

Then will their areas first, give, what

Here the area $ab$ is

Thus in this progression

Here the area $ab$ is

Also in this table, we may insert the figures, and the properties of

Suppose $a, b, c$ is a square

In this progression

The same may be done thus, of areas of $a, b, c,$ and we are in this progression

The same may be done thus, of areas of $a, b, c,$ and we are in this progression

Copyright © Cambridge University Library
Licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License (CC BY-NC 3.0)
In our study of the Gaussian distribution, we needed to compute integrals of the form
\[ \int_0^x e^{-t^2} \, dt \] (38)
However we compute the integral (38) using the Fundamental Theorem of Calculus, because the antiderivative of \( e^{-x^2} \) cannot be made out of the functions we know.

However, we can compute (38) using a completely new approach to functions and integration, that was discovered by Newton after he studied Wallis’ *Arithmetica Infinitorum*. On the facing page we see Newton’s handwritten notes from 1664 *Annotation out of Dr Wallis his Arithmetica Infinitorum*. \(^7\) In this section we will try to understand this page of Newton, in modern language.

Newton’s main point here is that we obtain more information if Wallis’ definite integrals
\[ \int_0^1 (1 - x^{1/p})^q \, dx \quad \text{(Wallis)} \]
are replaced by indefinite integrals
\[ \int_0^x (1 - t^{1/p})^q \, dt \quad \text{(Newton)}. \]
Following Wallis, Newton first computed these integrals when \( q \) is an integer, and then found a formula that made sense when \( q \) is not an integer, so he could use Wallis’ method of interpolation to guess the answer. This work is in the tables on the left side of the page.

Recall that for an integer \( n \geq 0 \) the binomial expansion is the algebraic formula
\[
(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.
\] (39)
Recall from homework that
\[
\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{1 \cdot 2 \cdot 3 \cdots k},
\]
so (39) can be written as
\[
(1 + x)^n = 1 + nx + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} x^2 + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} x^3 + \cdots + \frac{n(n-1)(n-2)\cdots 1}{1 \cdot 2 \cdot 3 \cdots n} x^n. \] (40)
If we kept going, the coefficients would be
\[ \frac{n(n-1)(n-2)\cdots 1 \cdot 0}{1 \cdot 2 \cdot 3 \cdots n \cdot (n+1)} = 0, \]
\[ \frac{n(n-1)(n-2)\cdots 1 \cdot 0 \cdot (-1)}{1 \cdot 2 \cdot 3 \cdots n \cdot (n+1) \cdot (n+2)} = 0, \quad \text{etc.} \]

\(^7\)The mathematical papers of Isaac Newton, vol. I. pp. 96-115. The image comes from http://cudl.lib.cam.ac.uk/view/MS-ADD-03958/139. Change the last number to see more.
So we can write it without the last term:

\[(1 + x)^n = 1 + nx + \frac{n(n - 1)}{1 \cdot 2} x^2 + \frac{n(n - 1)(n - 2)}{1 \cdot 2 \cdot 3} x^3 + \ldots \]

\[= \sum_{k=0}^{\infty} \frac{n(n - 1)(n - k + 1)}{1 \cdot 2 \ldots k} x^k, \tag{41} \]

Or more briefly,

\[(1 + x)^n = \sum_{k=0}^{\infty} \binom{n}{k} x^k. \tag{42} \]

Both sides now seem to make sense for any number \(n\). However if \(n\) is not a positive integer, the right side is an infinite sum. Newton guessed that (42) might still be true in this case, which would be extremely useful, because it is easy to integrate powers:

\[\int_0^x t^k \, dt = \frac{x^{k+1}}{k+1}. \tag{43} \]

So Newton worked out some examples. He was well aware that (42) had not been proved yet, so he qualified all of his results with “It may appear . . .”

**Newton’s first example: Quadrature of the Hyperbola**

Here Newton used his new method to compute the area under the hyperbola \(y = (1 + x)^{-1}\) from 0 to \(x\). Taking \(n = -1\) in (41), he found that

\[\frac{1}{1 + x} = 1 + (-1)x + \frac{(-1)(-2)}{1 \cdot 2} x^2 + \frac{(-1)(-2)(-3)}{1 \cdot 2 \cdot 3} x^3 + \ldots , \]

or,

\[\frac{1}{1 + x} = 1 - x + x^2 - x^3 + x^4 - x^5 + \ldots . \]

Replacing \(x\) by \(t\) and integrating, he found the area under the hyperbola “may appear” to be

\[\text{Area} = \int_0^x \frac{1}{1 + t} \, dt = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \frac{x^6}{6} + \ldots . \]

On the other hand, for the simple function \((1 + t)^{-1}\) we can use the FTC and we get

\[\int_0^x \frac{1}{1 + t} \, dt = \log(1 + x). \]

Comparing the two ways of computing the integral, it “may appear” that

\[\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \frac{x^6}{6} + \ldots . \]

You can see this in the top section of the page.

If true, this means the logarithm function is actually a polynomial that happens to have no highest degree term. Could it be that *every function* is such a polynomial? If so, then we could integrate all functions uniformly, without tricks, just using (43).
Newton’s Second example: Quadrature of the Circle

Recall Wallis found the area of a whole circle, in a roundabout way. Here Newton finds the area of an arbitrary segment of the circle, in a systematic way. Take $n = 1/2$ in (41) to get

$$\sqrt{1 + x} = \sum_{k=0}^{\infty} \left( \frac{1/2}{k} \right) x^k = 1 + \sum_{k=1}^{\infty} (-1)^{k-1} \frac{P_k}{2k + 1} x^k.$$  

The last equation used the result from our homework

$$\left( \frac{1/2}{k} \right) = (-1)^{k-1} \frac{P_k}{2k - 1}, \quad \text{where} \quad P_k = \frac{1 \cdot 3 \cdots (2k - 1)}{2 \cdot 4 \cdots (2k)}.$$  

Replacing $x$ by $-t^2$, we get

$$\sqrt{1 - t^2} = \sum_{k=0}^{\infty} \left( \frac{1/2}{k} \right) (-t^2)^k = 1 + \sum_{k=1}^{\infty} (-1)^{k-1} \frac{P_k}{2k - 1} (-t^2)^k = 1 - \sum_{k=1}^{\infty} \frac{P_k}{2k - 1} t^{2k}.$$  

Integrating, we get

$$\text{Area} = \int_0^x \sqrt{1 - t^2} \, dt = x - \sum_{k=1}^{\infty} \frac{P_k}{(2k - 1)(2k + 1)} x^{2k + 1}$$

$$= x - \frac{1}{6} x^3 - \frac{1}{40} x^5 - \frac{1}{112} x^7 - \frac{5}{1152} x^7 \cdots,$$

as we see written by Newton near the bottom of the middle section, ninth line from the end.

Newton’s Third example: Arcsine

Take $n = -1/2$ in (41) to get

$$\frac{1}{\sqrt{1 + x}} = \sum_{k=0}^{\infty} \left( \frac{-1/2}{k} \right) x^k = \sum_{k=0}^{\infty} (-1)^k P_k x^k.$$  

Replacing $x$ by $-t^2$ we get

$$\frac{1}{\sqrt{1 - t^2}} = \sum_{k=0}^{\infty} \left( \frac{-1/2}{k} \right) (-t^2)^k = \sum_{k=0}^{\infty} P_k t^{2k}.$$  

using the formula from homework:

$$\left( \frac{-1/2}{k} \right) = (-1)^k P_k;$$  

Integrating, we get

$$\int_0^x \frac{1}{\sqrt{1 - t^2}} \, dt = \sum_{k=0}^{\infty} P_k \frac{x^{2k+1}}{2k + 1}.$$  

On the other hand, we know that

$$\int_0^x \frac{1}{\sqrt{1 - t^2}} \, dt = \arcsin x.$$  

37
Thus, Newton finds

\[
\arcsin x = \sum_{k=0}^{\infty} P_k \frac{x^{2k+1}}{2k+1}
\]

\[
= x + \frac{1}{2} \cdot \frac{x^3}{3} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \cdot \frac{x^5}{5} + \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8} \cdot \frac{x^7}{7} + \cdots
\]

\[
= x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{5x^7}{112} + \frac{35x^9}{1152} + \cdots
\]

This series, multiplied by 1/2, appears five lines from the bottom of the page.

Thus, on this one page, Newton has written down infinite series for the functions \((1 + x)^n\) \((n\text{ any number})\), \(\log(1 + x)\) and \(\arcsin x\).

9 Binomial Theorem and the ellipse

Before Newton, Kepler had observed that planets move in elliptical orbits with the sun at one focus. How far is the orbit? In other words, what is the circumference, or arclength, of an ellipse?

An ellipse is a kind of flattened circle, where the amount of flattening is measured by the eccentricity, which literally means “out of round-ness”, and which we shall denote by \(\varepsilon\). It is defined as the ratio

\[
\varepsilon = \frac{\text{distance from the center to a focus}}{\text{length of semimajor axis}}.
\]

We can choose coordinates so that our ellipse has equation

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,
\]

where \(a \geq b\). Then the foci are \((\pm \sqrt{a^2 - b^2}, 0)\) and

\[
\varepsilon = \sqrt{1 - \frac{b^2}{a^2}}.
\]

Note that \(0 \leq \varepsilon \leq 1\). If \(\varepsilon = 0\) then \(a = b\) so the ellipse is a circle of radius \(a\). If we squash the ellipse vertically, by decreasing \(b\) and keeping \(a\) fixed, then \(\varepsilon\) increases from 0 to 1; when \(\varepsilon = 1\) then \(b = 0\), so the ellipse has been completely flattened into the major axis \([-a, a]\).

For example, the eccentricities of the six planets known to Kepler are
<table>
<thead>
<tr>
<th>Planet</th>
<th>ε</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury</td>
<td>.2056</td>
</tr>
<tr>
<td>Venus</td>
<td>.0068</td>
</tr>
<tr>
<td>Earth</td>
<td>.0167</td>
</tr>
<tr>
<td>Mars</td>
<td>.0934</td>
</tr>
<tr>
<td>Jupiter</td>
<td>.0484</td>
</tr>
<tr>
<td>Saturn</td>
<td>.0542</td>
</tr>
</tbody>
</table>

Imagine a planet moving on a general ellipse in the \((x, y)\) plane. The distance travelled is the integral of the planet’s speed. If \((x(t), y(t))\) is the position of the planet at time \(t\), then its velocity vector is \((x'(t), y'(t))\) and its speed is

\[
v(t) = \sqrt{x'(t)^2 + y'(t)^2}.
\]

Over a time interval \([t_1, t_2]\), the distance travelled by the planet is then

\[
\int_{t_1}^{t_2} v(t) \, dt = \int_{t_1}^{t_2} \sqrt{x'(t)^2 + y'(t)^2} \, dt. \tag{44}
\]

Suppose

\[
x(t) = a \cos t, \quad y(t) = b \sin t,
\]

so that the orbit is an ellipse with equation

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,
\]

traversed exactly once during the time interval \(0 \leq t \leq 2\pi\), and the speed is

\[
v(t) = \sqrt{a^2 \sin^2 t + b^2 \cos^2 t}.
\]

The quadrants divide the ellipse into four equal parts, so let us just measure our planet’s journey in the first quadrant, for \(0 \leq t \leq \pi/2\), and then multiply by 4. The length of this quarter-ellipse is then given by the **elliptic integral**

\[
L = \int_{0}^{\pi/2} \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} \, dt. \tag{45}
\]

If the ellipse were actually a circle, that is, if \(a\) were equal to \(b\), the elliptic integral would be simply

\[
\int_{0}^{\pi/2} \sqrt{a^2 \sin^2 t + a^2 \cos^2 t} \, dt = \int_{0}^{\pi/2} a \, dt = a \frac{\pi a}{2},
\]

which we know is the length of a quarter-circle.

But if \(a \neq b\), then the elliptic integral is another one of those hard integrals which cannot be evaluated by finding an antiderivative.
In his textbook *Calculus Integralis* Euler shows how to calculate the integral $L$ as a power series in the eccentricity $\varepsilon$. His method depends on the Wallis integrals $I_{2k}$ that we have studied, Wallis, and the coefficients of Euler’s power series are expressed in terms of Wallis’ fractions.  

We will use Newton’s binomial series. First note that $b^2 = (1 - \varepsilon^2)a^2$. Now we have

\[
a^2 \sin^2 t + b^2 \cos^2 t = a^2 (1 - \cos^2 t) + a^2 (1 - \varepsilon^2) \cos^2 t \]

\[
= a^2 (1 - \varepsilon^2 \cos^2 t),
\]

so we have

\[
L = a \int_0^{\pi/2} \sqrt{1 - \varepsilon^2 \cos^2 t} \, dt.
\]

We will expand the integrand using the binomial series. Recall that

\[
\left(\frac{1/2}{k}\right) = (-1)^{k-1} \frac{P_k}{2k-1},
\]

so

\[
\sqrt{1 + x} = \sum_{k=0}^{\infty} \left(\frac{1/2}{k}\right) x^k = \sum_{k=0}^{\infty} (-1)^{k-1} \frac{P_k}{2k-1} x^k = 1 - \sum_{k=1}^{\infty} (-1)^k \frac{P_k}{2k-1} x^k,
\]

where as always, $P_k = \frac{1 \cdot 3 \cdot \ldots \cdot (2k-1)}{2 \cdot 4 \cdot \ldots \cdot (2k)}$. Now substitute $-\varepsilon^2 \cos^2 t = x$. The $(-1)^k$’s cancel and we get

\[
\sqrt{1 - \varepsilon \cos^2 t} = 1 - \sum_{k=1}^{\infty} \frac{P_k}{2k-1} \varepsilon^{2k} \cos^{2k} t.
\]

Now we integrate to get

\[
L = a \left[ \int_0^{\pi/2} \left(1 - \sum_{k=1}^{\infty} \frac{P_k}{2k-1} \varepsilon^{2k} \cos^{2k} t \right) \right].
\]

Our basic integrals $I_{2k} = \int_0^{\pi/2} \cos^{2k} t$ have appeared once again, now in cosine form. Recall that

\[
\int_0^{\pi/2} \cos^{2k} t \, dt = \frac{\pi}{2} P_k.
\]

Putting this into the summation, we get

\[
L = \frac{\pi a}{2} \left[ 1 - \sum_{k=1}^{\infty} \frac{P_k^2}{2k-1} \varepsilon^{2k} \right].
\]

This is Euler’s formula for the arclength of the quarter-ellipse.

---

\(^8\)Euler actually uses a different parametrization of the ellipse (via rational functions) and his calculations are more complicated than what follows, but we’ll arrive at the same result he did.
The whole ellipse has circumference

\[ 4L = 2\pi a \cdot \left[ 1 - \sum_{k=1}^{\infty} \frac{P_k^2}{2k-1} \varepsilon^{2k} \right] \]

\[ = 2\pi a \cdot \left[ 1 - \left( \frac{1}{2} \right)^2 \cdot \varepsilon^2 - \left( \frac{1 \cdot 3}{2 \cdot 4} \right)^2 \cdot \varepsilon^4 - \left( \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \right)^2 \cdot \varepsilon^6 - \cdots \right] \]  \( (46) \)

Thus, the circumference of the ellipse is expressed as a power series in the eccentricity \( \varepsilon \).

Note that for \( \varepsilon = 0 \), the circle case, we get \( 4L = 2\pi a \), the circumference of a circle, as we should. If we fix \( a \) and flatten the circle by decreasing \( b \), so that \( \varepsilon \) increases from 0 to 1, the series measures the decrease in arclength as the circle is flattened. If we flatten all the way to \( b = 0 \), then \( \varepsilon = 1 \) and our quarter-ellipse is just the line segment \([0, a]\). So it appears that

\[ a = \frac{a\pi}{2} \left[ 1 - \sum_{k=1}^{\infty} \frac{P_k^2}{2k-1} \right], \]  \( (47) \)

provided the series converges. From the probabilistic version of Wallis’ formula (20), we have

\[ \frac{4k^2 - 1}{1} \cdot \frac{P_k^2}{2k-1} = (2k + 1)P_k^2 \to \frac{2}{\pi}. \]

Hence the series (47) converges by limit comparison with

\[ \sum \frac{1}{4k^2 - 1} \]

and gives the formula

\[ \frac{2}{\pi} = 1 - \sum_{k=1}^{\infty} \frac{P_k^2}{2k-1}, \]

which is a summation-version of Wallis’ product formula.

**Exercise** Some people use the simpler formula

\[ 2\pi \sqrt{a^2 + b^2} = 2\pi a \sqrt{1 - \frac{\varepsilon^2}{2}} \]  \( (48) \)

for the circumference of the ellipse.

a) Find a power series in \( \varepsilon \) for

\[ 2\pi a \sqrt{1 - \frac{\varepsilon^2}{2}} \]

b) Compare your power series in a) with the power series in (46). Why are they close? Which is bigger?
An idealized pendulum consists of a rod hanging from a fixed pivot with a weight on the lower end, swinging back and forth with no friction or air resistance. We also assume the mass hanging on the end is much greater than the mass of the rod itself, like the pendulum in a grandfather clock, so we can ignore the mass of the rod.

Let $\ell$ be the length of the rod, $m$ the mass on the end, $g$ the acceleration due to gravity, and $\theta(t)$ the radian angle of the rod with the vertical. After an initial displacement of $\theta_0 = \theta(0)$ radians, we let the mass go, so it swings freely back and forth. The period is the time $T$ that it takes the mass to return to its initial displacement.

In elementary physics, they use Newton’s law $F = ma$ to show that

$$\theta''(t) + \frac{g}{\ell} \sin \theta(t) = 0.$$  

This is a nonlinear differential equation (because $\theta(t)$ is inside the sine function) that is difficult to solve, so they replace $\sin(\theta(t))$ by $\theta(t)$, getting a linear differential equation

$$\theta''(t) + \frac{g}{\ell} \theta(t) = 0. \quad (49)$$

Now $\theta(0) = \theta_0$ and $\theta'(0) = 0$ because we just let the mass go. One can show that the only function satisfying the linear differential equation (49) and these initial conditions is

$$\theta(t) = \theta_0 \cdot \cos \left( t \sqrt{\frac{g}{\ell}} \right).$$

The rod is vertical when $\theta(t) = 0$, which first happens at time

$$t = \frac{\pi}{2} \cdot \sqrt{\frac{\ell}{g}}.$$  

The rod returns to its original position after four such time intervals, so we get the approximation

$$T \approx 2\pi \cdot \sqrt{\frac{\ell}{g}} \quad \text{(assuming } \sin \theta \approx \theta).$$

Using power series, we can give an exact formula for $T$. First we find another equation for $\theta$ using conservation of energy which says that the kinetic energy gained by falling equals the loss of potential energy at the starting point. This is expressed mathematically as

$$\frac{1}{2} m \cdot v(t)^2 = mg \cdot h(t) \quad (50)$$

$^9$This is not a bad approximation when $\theta(t)$ is small, since the Taylor series for $\sin \theta$ is $\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots$. 

42
where \( v(t) \) is the speed and \( h(t) \) is the distance fallen, both at time \( t \). We have \( v = \ell \theta' \) and \( h = \ell \cos \theta - \ell \cos \theta_0 \), so equation (50) becomes

\[
\frac{d\theta}{dt} = \sqrt{\frac{2g}{\ell}} \sqrt{\cos \theta - \cos \theta_0}.
\]

This describes \( \theta \) as a function of \( t \), but we want to know \( t \) as a function of \( \theta \), so we can evaluate this function at \( \theta_0 \). Now

\[
\frac{dt}{d\theta} = \sqrt{\frac{\ell}{2g}} \frac{1}{\sqrt{\cos \theta - \cos \theta_0}},
\]

so

\[
t = \sqrt{\frac{\ell}{2g}} \int \frac{d\theta}{\sqrt{\cos \theta - \cos \theta_0}},
\]

where the antiderivative is chosen to be zero when \( \theta = \theta_0 \), since that is the angle when \( t = 0 \).

Remembering that the mass must swing through the angle \( \theta_0 \) four times to get back to its starting point, we find the period is given by the integral

\[
T = 4 \sqrt{\frac{\ell}{2g}} \int_0^{\theta_0} \frac{d\theta}{\sqrt{\cos \theta - \cos \theta_0}}.
\]

We now make some elementary-but-clever manipulations and substitutions to make this integral more familiar. Using the identity \( \cos 2\alpha = 1 - 2\sin^2 \alpha \), we have

\[
\begin{align*}
\cos \theta - \cos \theta_0 &= \cos 2(\theta/2) - \cos 2(\theta_0/2) \\
&= 2(\sin^2(\theta_0/2) - \sin^2(\theta/2)) \\
&= 2\sigma^2(1 - \sin^2 u) \\
&= 2\sigma^2 \cos^2 u,
\end{align*}
\]

where

\[
\sin u = \frac{\sin(\theta/2)}{\sin(\theta_0/2)} \quad \text{and} \quad \sigma = \sin(\theta_0/2),
\]

so \( \sigma \sin u = \sin(\theta/2) \). Converting the integration variable to \( u \), we have

\[
\sigma \cos u \, du = \frac{1}{2} \cos(\theta/2) \, d\theta = \frac{1}{2} \sqrt{1 - \sin^2(\theta/2)} \, d\theta = \frac{1}{2} \sqrt{1 - \sigma^2 \sin^2 u} \, d\theta,
\]

so

\[
d\theta = \frac{2\sigma \cos u}{\sqrt{1 - \sigma^2 \sin^2 u}} \, du.
\]

Putting everything together, we have

\[
T = 4 \sqrt{\frac{\ell}{g}} \int_0^{\pi/2} \frac{du}{\sqrt{1 - \sigma^2 \sin^2 u}}.
\]
This is very similar to the arclength integral for the ellipse. The main differences are that the square-root is in the denominator and the eccentricity $\varepsilon$ is now replaced by $\sigma = \sin(\theta_0/2)$, reflecting the initial displacement of the rod.

The same method will now express $T$ as a power series in $\sigma$. You will find that the constant term of this power series is $2\pi \sqrt{\ell/g}$, which is the approximation obtained above.