The Most Prestigious Family and Their Most Illustrious Pupil together with the most underappreciated mathematician of the period (if not of all time)

As we enter the eighteenth century, we see an explosion of mathematical power and knowledge. Calculus is being applied in unexpected ways, in almost any branch of the physical sciences, and with great success; and statistics is becoming a new way to achieve knowledge. So pervasive in our century, statistics is hardly a tool until the end of this period. We will study the most prestigious family in mathematical history, the Bernoulli’s. We will look into three members of the Swiss family, the founding brothers, Jacob (or Jacques or James) and Johann (or Jean or John), and the latter’s most prominent son, Daniel, and the best mathematician of the century, another Swiss, a student of Johann Bernoulli and a friend of Daniel Bernoulli, Euler. We also look at DeMoivre, who should be better known than he is.

James & His Glorious Theorem

An old Arab saying proclaims: Indeed he knows not how to know who knows not also how to un-know. James (1654-1705) made fundamental progress in how to test our knowledge, and in doing so became one of the first statisticians in history. He taught at the University of Basel, the same city where Euler is going to be born. Although his mathematical interests were very wide (typical of the age), his most lasting influence was in the field of probability and statistics. In 1713, eight years after his death, his nephew Nicholas (son of John) published, posthumously, Jacob’s masterpiece: Ars Conjectandi, an important book, the heir to Huygens’, and the predecessor to Laplace’s. The following discussion is an interpretation of some of the ideas Bernoulli discussed.

Suppose you have an urn in which you can hear some balls inside. Unfortunately the balls are too large to get them out, and certainly the urn is to be preserved, so you don’t want to destroy it. However through the opening you can see one of the balls (one at-a-time) inside, and if you rattle the urn you can perhaps change the ball you are seeing. Indeed, after a little while, through the top you see balls of two colors: black and white. You spend some time observing the different colors of the balls you see from the top of the urn, and you are ready to guess that perhaps there are five balls inside the urn: 3 black and 2 white. But you are an honest person with integrity and you would like some moral certainty concerning your guess. How would you go about it? How would we test our knowledge? In Bernoulli’s own words:

...how often a white and how often a black pebble is observed. The question is, can you do this so often that it becomes ten times, one hundred times, one thousand times, etc. more probable (that is, it be morally certain) that the numbers of whites and blacks observed (chosen) are in the same 3:2 ratio as the pebbles in the urn, rather than in any other ratio?

At the same time, Bernoulli realized that we could only expect an approximation to our
ratio, not an exact ratio. Namely, the more times we did the experiment, the less likely would it be that we get the exact ratio. Indeed, let's do some computations. Suppose we observe the urn 500 times, and suppose that our hypothesis of 3 black and 2 white is correct. Then the probability that we observe 300 black and 200 white is exactly:

\[
P(500,300) = \binom{500}{300} \left( \frac{3}{5} \right)^{300} \left( \frac{2}{5} \right)^{200} = \frac{500!3^{300}2^{200}}{300!200!5^{500}}.
\]

In general, Bernoulli knew that if his hypothesis was correct, and he did the experiment \(n\) times, the probability that he would get exactly \(k\) black showings was:

\[
P(n,k) = \binom{n}{k} \left( \frac{3}{5} \right)^k \left( \frac{2}{5} \right)^{n-k} = \frac{n!3^k2^{n-k}}{k!(n-k)!5^n}.
\]

This is one instance of what is often referred to as the binomial distribution. He understood enough about these numbers to appreciate that the larger the \(n\) got the smaller this number got regardless of what \(k\) was. Indeed he said

> To avoid misunderstanding, we must note that the ratio between the number of cases, which we are trying to determine by experiment, should not be taken as precise and indivisible (for then just the contrary would happen, and it would become less probable that the true ratio would be found the more numerous were the observations). Rather, it is the ratio taken with some latitude, that is, included within two limits, which can be made as narrow as one might wish. For instance, if in the example...above we take two ratios 301/200 and 299/200, or 3001/2000 and 2999/2000, etc., of which one is immediately greater and the other immediately less than the ratio 3:2, it will be shown that it can be made more probable, that the ratio found by often repeated experiments will fall within these limits of the 3:2 ratio rather than outside them.

And if we peek ahead at the table we see that as the number of experiments increase, the probability that we get exactly the ratio 3:2 decreases:

<table>
<thead>
<tr>
<th>Number of Trials</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.11456</td>
</tr>
<tr>
<td>100</td>
<td>0.08122</td>
</tr>
<tr>
<td>500</td>
<td>0.04069</td>
</tr>
</tbody>
</table>

as we go from 50 experiments to 500 by increments of 50 each. Thus for the case of 500, we get around 3.6% chance of coming up with a 3:2 ratio.

The table below represents a table of probabilities for different number of trials as indicated by the columns going from 50 to 500. The row labeled 0 indicates the probability of the exact ratio 3:2. Thus in 50 tries, it indicates 30 blacks versus 20 whites, while in the 100 column, it indicates 60 blacks versus 40 whites, etc. The respective rows indicate that many trials away from that 3:2 ratio.
<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>350</th>
<th>400</th>
<th>450</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.00199</td>
<td>0.01034</td>
<td>0.01660</td>
<td>0.02023</td>
<td>0.02225</td>
<td>0.02385</td>
<td>0.02403</td>
<td>0.02402</td>
<td>0.02387</td>
<td></td>
</tr>
<tr>
<td>-9</td>
<td>0.00426</td>
<td>0.01520</td>
<td>0.02152</td>
<td>0.02461</td>
<td>0.02604</td>
<td>0.02660</td>
<td>0.02669</td>
<td>0.02653</td>
<td>0.02622</td>
<td>0.02584</td>
</tr>
<tr>
<td>-8</td>
<td>0.00842</td>
<td>0.02149</td>
<td>0.02716</td>
<td>0.02933</td>
<td>0.02998</td>
<td>0.02992</td>
<td>0.02954</td>
<td>0.02899</td>
<td>0.02838</td>
<td>0.02774</td>
</tr>
<tr>
<td>-7</td>
<td>0.01537</td>
<td>0.02919</td>
<td>0.03337</td>
<td>0.03426</td>
<td>0.03396</td>
<td>0.03321</td>
<td>0.03230</td>
<td>0.03135</td>
<td>0.03043</td>
<td>0.02954</td>
</tr>
<tr>
<td>-6</td>
<td>0.02594</td>
<td>0.03811</td>
<td>0.03993</td>
<td>0.03972</td>
<td>0.03785</td>
<td>0.03636</td>
<td>0.03491</td>
<td>0.03357</td>
<td>0.03233</td>
<td>0.03120</td>
</tr>
<tr>
<td>-5</td>
<td>0.04046</td>
<td>0.04781</td>
<td>0.04651</td>
<td>0.04400</td>
<td>0.04151</td>
<td>0.03927</td>
<td>0.03730</td>
<td>0.03556</td>
<td>0.03404</td>
<td>0.03268</td>
</tr>
<tr>
<td>-4</td>
<td>0.05836</td>
<td>0.05763</td>
<td>0.05273</td>
<td>0.04836</td>
<td>0.04478</td>
<td>0.04184</td>
<td>0.03938</td>
<td>0.03730</td>
<td>0.03551</td>
<td>0.03395</td>
</tr>
<tr>
<td>-3</td>
<td>0.07781</td>
<td>0.06673</td>
<td>0.05818</td>
<td>0.05208</td>
<td>0.04752</td>
<td>0.04396</td>
<td>0.04109</td>
<td>0.03871</td>
<td>0.03670</td>
<td>0.03498</td>
</tr>
<tr>
<td>-2</td>
<td>0.09588</td>
<td>0.07421</td>
<td>0.06248</td>
<td>0.05495</td>
<td>0.04961</td>
<td>0.04557</td>
<td>0.04238</td>
<td>0.03977</td>
<td>0.03759</td>
<td>0.03574</td>
</tr>
<tr>
<td>-1</td>
<td>0.10910</td>
<td>0.07924</td>
<td>0.06529</td>
<td>0.05680</td>
<td>0.05094</td>
<td>0.04659</td>
<td>0.04319</td>
<td>0.04044</td>
<td>0.03815</td>
<td>0.03622</td>
</tr>
<tr>
<td>0</td>
<td>0.11456</td>
<td>0.08122</td>
<td>0.06637</td>
<td>0.05751</td>
<td>0.05145</td>
<td>0.04697</td>
<td>0.04350</td>
<td>0.04069</td>
<td>0.03837</td>
<td>0.03640</td>
</tr>
<tr>
<td>1</td>
<td>0.11066</td>
<td>0.07989</td>
<td>0.06564</td>
<td>0.05703</td>
<td>0.05111</td>
<td>0.04671</td>
<td>0.04329</td>
<td>0.04052</td>
<td>0.03822</td>
<td>0.03628</td>
</tr>
<tr>
<td>2</td>
<td>0.09874</td>
<td>0.07538</td>
<td>0.06315</td>
<td>0.05540</td>
<td>0.04993</td>
<td>0.04582</td>
<td>0.04257</td>
<td>0.03994</td>
<td>0.03773</td>
<td>0.03586</td>
</tr>
<tr>
<td>3</td>
<td>0.08079</td>
<td>0.06820</td>
<td>0.05907</td>
<td>0.05269</td>
<td>0.04797</td>
<td>0.04431</td>
<td>0.04138</td>
<td>0.03895</td>
<td>0.03690</td>
<td>0.03515</td>
</tr>
<tr>
<td>4</td>
<td>0.06059</td>
<td>0.05914</td>
<td>0.05373</td>
<td>0.04908</td>
<td>0.04533</td>
<td>0.04227</td>
<td>0.03973</td>
<td>0.03759</td>
<td>0.03576</td>
<td>0.03416</td>
</tr>
<tr>
<td>5</td>
<td>0.04155</td>
<td>0.04913</td>
<td>0.04751</td>
<td>0.04476</td>
<td>0.04211</td>
<td>0.03975</td>
<td>0.03770</td>
<td>0.03590</td>
<td>0.03433</td>
<td>0.03293</td>
</tr>
<tr>
<td>6</td>
<td>0.02597</td>
<td>0.03908</td>
<td>0.04083</td>
<td>0.03997</td>
<td>0.03846</td>
<td>0.03687</td>
<td>0.03534</td>
<td>0.03393</td>
<td>0.03265</td>
<td>0.03148</td>
</tr>
<tr>
<td>7</td>
<td>0.01474</td>
<td>0.02975</td>
<td>0.03409</td>
<td>0.03493</td>
<td>0.03454</td>
<td>0.03371</td>
<td>0.03274</td>
<td>0.03174</td>
<td>0.03076</td>
<td>0.02984</td>
</tr>
<tr>
<td>8</td>
<td>0.00756</td>
<td>0.02166</td>
<td>0.02766</td>
<td>0.02988</td>
<td>0.03050</td>
<td>0.03040</td>
<td>0.02996</td>
<td>0.02937</td>
<td>0.02872</td>
<td>0.02805</td>
</tr>
<tr>
<td>9</td>
<td>0.00349</td>
<td>0.01507</td>
<td>0.02179</td>
<td>0.02502</td>
<td>0.02647</td>
<td>0.02702</td>
<td>0.02709</td>
<td>0.02689</td>
<td>0.02655</td>
<td>0.02614</td>
</tr>
<tr>
<td>10</td>
<td>0.00144</td>
<td>0.01001</td>
<td>0.01667</td>
<td>0.02050</td>
<td>0.02258</td>
<td>0.02368</td>
<td>0.02419</td>
<td>0.02436</td>
<td>0.02433</td>
<td>0.02416</td>
</tr>
</tbody>
</table>

The table above possesses some very meaningful patterns. If we look at any one column we see that the highest probability does in fact occur at the expected ratio 3:2 (as it should be since the highest binomial coefficient always occurs in the middle). But as the number of trials increases, this probability decreases—as we pointed out above—since now the tallest member has a lot of close competitors. The picture makes any more words unnecessary.

Bernoulli then made a wonderful observation. If we give up on the exactness of the ratio but we are content with staying within a fixed fraction of that ratio, then we can indeed expect our probabilities to increase as the number of experiments increases. Let us say, as he did, that we want to stay within \( \frac{1}{50} \) (2%) of our fixed ratio.

Thus if we do the experiment 50 times, then we expect to have either 29, 30 or 31 successes (black balls), if we do it 100 times, then we look at the probability of 58, 59, 60, 61 or 62 successes, etcetera.

The table below gives both the number of occurrences of black balls and the probabilities for each which have been added up at the bottom of the table. The top row indicates the number of experiments.
And we see the probabilities climb as the number of experiments increase:

This fact, namely, that the probability of staying within an arbitrary, yet prescribed, fraction from the exact ratio does increase without bound all the way to 1 as the number of experiments increases without bound is known sometimes as the (weak) Law of Large Numbers, or as the Law of Averages, or sometimes as Bernoulli’s Theorem. And we spend the remainder of this section visiting Bernoulli’s proof of his theorem. In order to prove it, Bernoulli needed four ingredients. The first one of which is a very pretty inequality:

**Lemma 1.** Let \( a, b, c, d \) be positive numbers. If \( \frac{a}{b} < \frac{c}{d} \), then \( \frac{a + c}{b + d} \) is in between the other two, \( \frac{a}{b} < \frac{a + c}{b + d} < \frac{c}{d} \).

**Proof.** Since \( \frac{a}{b} < \frac{c}{d} \), we have \( ad < bc \). Hence \( ab + ad < ab + bc \) which makes \( \frac{a}{b} < \frac{a + c}{b + d} \) true. The other inequality is similarly obtained.

We can not help observing that this is an average that does occur in baseball. If a player goes \( \frac{3}{5} \) (it is customarily read: 3 for 5) one day and \( \frac{4}{6} \) another, then her/his combined total for the two days is \( \frac{7}{11} \), which is necessarily in between—not as good as her/his best
day and not as bad as her/his worse day. Observe that although \( \frac{4}{6} \) is the same as \( \frac{2}{3} \) in most contexts, it is not so from this point of view, since \( \frac{5}{8} \) (which is what we get when we average \( \frac{3}{5} \) and \( \frac{2}{3} \)) is not the same as \( \frac{7}{11} \).

From the inequality, it easily follows that for any positive \( k \), if \( a_1, a_2, \ldots, a_k \) and \( b_1, b_2, \ldots, b_k \) are positive numbers, and
\[
\frac{a_1}{b_1} < \frac{a_2}{b_2} < \cdots < \frac{a_k}{b_k}
\]
then
\[
\frac{a_1}{b_1} < \frac{a_1 + a_2 + \cdots + a_k}{b_1 + b_2 + \cdots + b_k} < \frac{a_k}{b_k}
\]

**Lemma 2** is also very well known and very important. It concerns the most important series, the geometric series, and although we have seen it before, we will remind ourselves.

**Lemma 2.** Let \( \beta \) be a positive number less than 1. Then
\[
\beta + \beta^2 + \beta^3 + \cdots = \frac{\beta}{1-\beta}.
\]

**Proof.** By simple multiplication, for any \( k \), \( \beta + \beta^2 + \beta^3 + \cdots + \beta^k = \frac{\beta - \beta^{k+1}}{1-\beta} \). Since \( \beta < 1 \), as \( k \) grows very large, \( \beta^{k+1} \to 0 \).

Suppose now we let \( N \) be the **total number of experiments**, and because we chose the fraction \( \frac{1}{50} \), we are going to let \( N = 50M \). Then the ratio 3:2 occurs when we have \( 30M \) observations of black balls. Hence to be within \( \frac{1}{50} \) of that ratio means we must have between \( 29M \) and \( 31M \) occurrences of black balls (endpoints included). We are interested then in
\[
\Omega = \text{probability that we have the number of occurrences of black balls fall within these two limits},
\]
and we would like to prove that as \( n \) grows, so does \( \Omega \), and that \( \frac{\Omega}{1-\Omega} \) becomes arbitrarily large (moral certainty).

We will let \( X \) denote the number of black ball occurrences. Then the next ingredient is fairly easy too:

**Lemma 3.** Let \( a \) and \( b \) be integers and let \( a > b \). Then
\[
\frac{P(X = a)}{P(X = b)} > \frac{P(X = a + 1)}{P(X = b + 1)}.
\]

**Proof.** From (*) above we know the left hand side of the inequality is given by
\[ P(X = a) = \frac{N!3^a 2^{N-a}}{a!(N-a)!5^N} \]
\[ P(X = b) = \frac{a!(N-a)!5^N}{N!3^b 2^{N-b}} = \frac{3^a 2^{N-a} b!(N-b)!}{3^b 2^{N-b} a!(N-a)!} \]

and the right hand side:

\[ P(X = a + 1) = \frac{N!3^{a+1} 2^{N-a-1}}{(a+1)!((N-a-1)!5^N) \frac{3^{a+1} 2^{N-a-1} (b+1)!((N-b-1)!}{3^b 2^{N-b-1} (a+1)!((N-a-1)!} \]

Hence if we compute \( \frac{P(X = b)}{P(X = a + 1)} \), we get

\[ \frac{3^a 2^{N-a} b!(N-b)3^{b+1} 2^{N-b-1}(a+1)!((N-a-1)!}{3^b 2^{N-b} a!(N-a)3^{a+1} 2^{N-a-1}(b+1)!((N-b-1)!} = (N-b)(a+1) \]

which is bigger than 1 since \( a+1 > b+1 \) and \( N-b > N-a \).

Finally, one more ingredient

**Lemma 4.** Let \( \beta = \frac{P(X = 31M)}{P(X = 30M)} \). Then \( \beta \to 0 \) as \( M \to \infty \).

**Proof.** From (*) above

\[ \beta = \frac{P(X = 31M)}{P(X = 30M)} = \frac{(31M)!31M!2^{19M}}{(30M)!30M!2^{20M}} \]

We would like to argue that \( \beta \leq \left( \frac{30}{31} \right)^n \)—and this would give us the lemma. The reasons for this are as follows:

\[ \beta = \frac{3^M (30M)!(20M)!}{2^M (31M)!(19M)!} = \frac{3^M (19M + 1)(19M + 2) \cdots (20M)}{2^M (30M + 1)(30M + 2) \cdots (31M)} \]

which is obtained by canceling some of the terms. Note that there are \( M \) factors in the denominator all of the form \( 30M + i \) (where \( i \leq M \)) and there are also \( M \) factors in the numerator all of the form \( 19M + i \). Any factor \( \frac{19M + i}{30M + i} \leq \frac{2(30)}{3(31)} \) since when we cross multiply we get \( 1767M + 93i \leq 1800M + 60i \) which is true since \( i \leq M \).

We can now prove the theorem. Recall that what we want to prove is that \( \frac{\Omega}{1-\Omega} \) grows without bound as \( M \to \infty \) (or equivalently, \( N \to \infty \)).
Let $A_0$ denote the probability that the number of occurrences, $X$, falls between $30M$ (exclusive) and $31M$ (inclusive), in symbols, $A_0 = P(30M < X \leq 31M)$.

Similarly, we define $A_1 = P(31M < X \leq 32M)$, $A_2 = P(32M < X \leq 33M)$, ..., $A_{19} = P(49M < X \leq 50M)$. Dually, we let, $B_0 = P(29M \leq X < 30M)$, $B_1 = P(28M \leq X < 29M)$, $B_2 = P(27M \leq X < 28M)$, ..., $B_{29} = P(0 \leq X < M)$.

Then we know that

$$\Omega > A_0 + B_0$$

(note we have not considered the case of exactly $30M$, and that is why we have strict inequality) and also that

$$1 - \Omega = A_1 + A_2 + \ldots + A_{19} + B_1 + B_2 + \ldots + B_{29}.$$ 

Hence if we can get $\frac{A_0}{A_1 + A_2 + \ldots + A_{19}}$ arbitrarily large (or equivalently, its reciprocal, arbitrarily small), and, if we prove a similar claim for the $B$’s, we will be done. But, observe

$$\frac{A_{k+1}}{A_k} = \frac{P(X = (30 + k + 1)M + 1) + P(X = (30 + k + 1)M + 2) + \ldots + P(X = (30 + k + 1)M + M)}{P(X = (30 + k)M + 1) + P(X = (30 + k)M + 2) + \ldots + P(X = (30 + k)M + M)}$$

which by Lemma 1 and Lemma 3, gives

$$\frac{A_{k+1}}{A_k} \leq \frac{P(X = (30 + k + 1)M + 1)}{P(X = (30 + k)M + 1)}.$$ 

By Lemma 3 again we get

$$\frac{A_{k+1}}{A_k} \leq \beta.$$ 

Hence $A_{k+1} \leq \beta A_k$ and thus $A_1 \leq \beta A_0$ and $A_2 \leq \beta A_1$, so $A_2 \leq \beta^2 A_0$, and similarly $A_3 \leq \beta^3 A_0$, $A_4 \leq \beta^4 A_0$, etcetera. Thus

$$\frac{A_1 + A_2 + \ldots + A_{19}}{A_0} \leq \beta + \beta^2 + \beta^3 + \ldots + \beta^{19} \leq \frac{\beta}{1 - \beta}$$

which gets arbitrarily small by Lemma 4. And we have proven Bernoulli’s Theorem.

But Bernoulli wanted actual estimates for the number of tries necessary, an estimate for $N$, and here perhaps he felt he had failed. For example, he wanted $\frac{\Omega}{1 - \Omega} \geq 1,000$, and then his estimate was 25,550 observations, which to him was a gigantic number—there were fewer than 3,000 known stars in the skies in his lifetime. De Moivre’s approximation to the binomial is an assault on improving $N$, but that will wait into later in this chapter.

Now we discuss his brother
The name of the younger John Bernoulli (1667-1748) would be even better known (than already is) if the theorem he provided L'Hospital to include in the latter's calculus book had received the proper reference of the creator of the theorem as opposed to the writer of the textbook.

As with his brother, his interests were very wide, but we will choose two topics that were close to his heart (and his brother’s) to concentrate on. The first one is a typical application of differential equations.

At $t = 0$, a rabbit starts running up the y-axis at constant speed $a$. Simultaneously, a dog starts chasing the rabbit starting at the point $(1,0)$ at constant speed $b$. The problem is to describe the path of the dog.

We will let $(x(t), y(t))$ describe the path of the dog. We certainly understand the path of the rabbit, at any time $t$, the rabbit is at the point $(0,at)$. Since at any time, the dog is aiming at the rabbit, the tangent line to $(x(t), y(t))$ passes through the point $(0,at)$, and so we have that

$$\frac{dy}{dx} = \frac{y(t) - at}{x(t)} \quad \text{(1)}$$

But we also know that $\frac{ds}{dt} = b$ where $s$ denotes the distance traveled by the dog. From calculus, we know that $ds^2 = dx^2 + dy^2$, and so

$$\frac{ds}{dx} = \pm \sqrt{1 + \left(\frac{dy}{dx}\right)^2} ,$$

and thus

$$\frac{dt}{dx} = \frac{dt}{ds} \frac{ds}{dx} = \frac{1}{b} \left( \pm \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \right) .$$

Since as $s$ grows, $x$ gets smaller, we know that $\frac{ds}{dx} < 0$, and so we can decide that the sign is negative. Hence we arrive at

$$\frac{dt}{dx} = -\frac{1}{b} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} .$$

In order to eliminate the $t$ then, we differentiate (1) with respect to $x$. More accurately, we
take the second derivative of \( x(t) \frac{dy}{dx} = y(t) - at \), to obtain
\[
x y'' + y' = y' - a \frac{ds}{dt} = y' + \frac{a}{b} \sqrt{1 + \left(y'\right)^2}.
\]
Although this is a second order differential equation, if we let \( z = y' \), then we have
\[
xz' + z = z + k \sqrt{1 + z^2}
\]
where \( k = \frac{a}{b} \). Canceling and separating:
\[
\frac{dz}{\sqrt{1 + z^2}} = kdx.
\]
Integrating,
\[
\ln \left( \sqrt{1 + z^2} + z \right) = \ln(x^k) + c.
\]
Now, when \( t = 0 \), \( x = 1 \), and \( z = 0 \) since the dog is aiming at the rabbit. Substituting, we have \( c = 0 \). So,
\[
\ln \left( \sqrt{1 + z^2} + z \right) = \ln(x^k)
\]
and thus,
\[
\sqrt{1 + z^2} + z = x^k.
\]
It follows,
\[
\sqrt{1 + z^2} = x^k - z.
\]
Squaring,
\[
1 + z^2 = x^{2k} - 2x^k z + z^2,
\]
and so
\[
z = x^k - \frac{x^{-k}}{2}.
\]
Integrating, provided \( k \neq 1 \), we get
\[
y = \frac{x^{k+1}}{2(k + 1)} - \frac{x^{-k+1}}{2(-k + 1)} + c
\]
while if \( k = 1 \), we get
\[
y = \frac{x^2}{4} - \ln \frac{x}{2} + c.
\]
Since, \( y = 0 \) when \( x = 1 \), we get
\[
y = \frac{x^{k+1}}{2(k + 1)} - \frac{x^{-k+1}}{2(-k + 1)} - \frac{k}{(k^2 - 1)} \quad \text{if} \quad k \neq 1
\]
while
\[
y = \frac{x^2 - 1}{4} - \ln \frac{x}{2} \quad \text{if} \quad k = 1.
\]
The graphs for some of the functions are:
Below is a table of the points on the y-axis where the dog will intersect the rabbit if \( k < 1 \).

<table>
<thead>
<tr>
<th>( k )</th>
<th>Intersection</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>10/99</td>
</tr>
<tr>
<td>0.2</td>
<td>5/24</td>
</tr>
<tr>
<td>0.3</td>
<td>30/91</td>
</tr>
<tr>
<td>0.4</td>
<td>10/21</td>
</tr>
<tr>
<td>0.5</td>
<td>2/3</td>
</tr>
<tr>
<td>0.6</td>
<td>15/16</td>
</tr>
<tr>
<td>0.7</td>
<td>1 19/51</td>
</tr>
<tr>
<td>0.8</td>
<td>2 2/9</td>
</tr>
<tr>
<td>0.9</td>
<td>4 14/19</td>
</tr>
</tbody>
</table>

We notice the increase as \( k \to 1 \).

Consider a particle falling from one point to another on a prescribed path, ignoring friction and all other forces except gravity. How much time will it take?

We will treat the question from the conservation of energy point of view and come to some of the same conclusions Bernoulli came to. Namely, we will exploit the fact that kinetic energy (given by \( \frac{1}{2}mv^2 \)) added to potential energy (given by \( gmh \)) will remain constant throughout the fall:

\[
\frac{1}{2}mv^2 + gmh = \text{constant,}
\]

where \( g \) is the gravitational constant, \( v \) is velocity and \( h \) denotes the height of the object. Clearly, \( v \) and \( h \) are variables depending on time \( t \).
In order to considerably simplify our calculations, we will consider an object of mass one \((m = 1)\) being dropped (so \(v(0) = 0\)) at \(t = 0\) at the point \((0,1)\) and timing its fall to the point \((1,0)\). Hence at all times \(h(t) = y(t)\), its \(y\)-coordinate.

Since \(v(0) = 0\) and \(h(0) = y(0) = 1\), by substituting \(t = 0\) in \((1)\), we can evaluate our constant to be \(g\). And so our equation \((1)\) above becomes:

\[v^2 = 2g(1 - y)\]

We know, \(v = \frac{ds}{dt}\) where \(s(t)\) indicates the distance traveled by time \(t\). Thus, we have from \((2)\), by taking square roots,

\[\frac{ds}{dt} = \sqrt{2g \sqrt{1 - y(t)}}\]

Bernoulli was interested in finding the path that would take the shortest time, and this is known as the **Brachistochrone Problem** (minimum time). We can not tackle the complete question with our limited tools, but we can demonstrate that the ***shortest path—the straight line—is not the quickest***. Additionally, we can hint at the complete solution.

In fact, we will start by comparing two specific paths, the **straight path**, and a **circular path** and show the circular path takes less time. The paths have respective equations:

\[x(t) + y(t) = 1\]
\[(x(t) - 1)^2 + (y(t) - 1)^2 = 1\].

In the case of a straight line, we have that, at all times \(t\), by the distance formula,

\[s(t) = \sqrt{x^2(t) + (1 - y(t))^2} = \sqrt{2(1 - y(t))}\],

since \(x(t) = 1 - y(t)\). Thus,

\[\sqrt{1 - y(t)} = \frac{s(t)}{\sqrt{2}}\]

and substituting into \((3)\), we obtain

\[\frac{ds}{dt} = \sqrt{2g \frac{s(t)}{\sqrt{2}}}\]

and after separating variables, we get the differential equation:

\[\frac{ds}{\sqrt{s}} = \sqrt{2g} dt\],

which can be solved by integrating both sides, becoming:

\[2\sqrt{s} = \sqrt{2} \sqrt{gt}\]

since \(s(0) = 0\), the constant of integration is 0. Hence the particle hits the floor when \(s = \sqrt{2}\), and hence at
For the circular path, we get \[ t = \frac{2}{\sqrt{g}}. \]

For the circular path, we get \[ \sin(s(t)) = 1 - y(t), \]

so becomes \[ \frac{ds}{dt} = \sqrt{2g} \sqrt{\sin(s)}, \]

and we can separate the variables again to obtain \[ \frac{ds}{\sqrt{\sin(s)}} = \sqrt{2g} dt. \]

Since we will be at the ground when \( s = \frac{\pi}{2} \), and we start when \( t = 0 \) and \( s = 0 \), we have that we will hit the ground when

\[
\int_0^\frac{\pi}{2} \frac{ds}{\sqrt{\sin(s)}} = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{4}\right)}{2\Gamma\left(\frac{3}{4}\right)}.
\]

Hence our time will approximately equal,

\[
t \approx \frac{2.622057554}{\sqrt{2g}} \approx \frac{1.854074677}{\sqrt{g}}.
\]

and thus it takes less time in the longer circular path.

But we can go further. The answer to the brachistochrone problem turns out to be the fascinating cycloid. In John’s own words:

*With justice we admire Huygens because he first discovered that a heavy particle falls on a cycloid in the same time always, no matter what the starting point may be. But you will be petrified with astonishment when I say that exactly this same cycloid, the tautochrone of Huygens, is the brachistochrone which we are seeking.*

To prove that the cycloid is the shortest path among all possible paths is a very different kind of problem, and new machinery, called the calculus of variations, was developed to tackled this kind of problem. Thus, we cannot accomplished our ultimate goal. However, we can and we will show that the cycloid is better than the circle. There are
many cycloids going through our two points, (1,0) and (0,1). Without giving the details, since we are dropping the particle \( v(0) = 0 \) at the point (0,1), we need a cycloid where (0,1) is at the highest point, and then there is only one such cycloid. And this is the one which will minimize time.

The path of this cycloid is given by:

\[
x(\theta) = a(\theta - \sin(\theta)), \\
y(\theta) = a(\cos(\theta) - 1) + 1
\]

where \( a \approx 0.572917037532 \). Note these are functions of \( \theta \), not \( t \). By elementary differential equations techniques, we get to the relation:

\[(s - 4a)^2 = 8a(y + 2a - 1).\]

Expanding and simplifying, we can solve for \( y \):

\[y = \frac{s^2}{8a} - s + 1.\]

Since we want to stop when \( y = 0 \), we easily find that \( s \approx 1.474106771 \) at that moment.

To find the time, we use 4 to first get \( \sqrt{1 - \frac{y}{s}} = \frac{\sqrt{8as - s^2}}{2\sqrt{2a}} \), and thus 3 becomes:

\[
\frac{ds}{dt} = \frac{\sqrt{g}}{2\sqrt{a}} \sqrt{8as - s^2},
\]

and separating variables one more time and integrating:

\[
\arcsin\left(\frac{s}{4a} - 1\right) = \frac{\sqrt{g}}{2\sqrt{a}} t + \text{constant},
\]

and since \( s(0) = 0 \), we evaluate the constant, and get:

\[
\arcsin\left(\frac{s}{4a} - 1\right) = \frac{\sqrt{g}}{2\sqrt{a}} t - \frac{\pi}{2}.
\]

By substituting \( s = 1.474106771 \), we finally compute:

\[
t = \frac{1.825682191}{\sqrt{g}},
\]

a shorter time than the circular path indeed! In order to simplify our summary of the results, we will assume that units of measurement have been arranged so that \( g = 1 \), and then as we travel from (1,0) to (0,1), we have:

<table>
<thead>
<tr>
<th>Path Chosen</th>
<th>Approximate Distance Traveled</th>
<th>Approximate Time Needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight Line</td>
<td>( \sqrt{2} \approx 1.414213562 )</td>
<td>2</td>
</tr>
<tr>
<td>Circular Path</td>
<td>( \frac{\pi}{2} \approx 1.570796327 )</td>
<td>1.854074677</td>
</tr>
<tr>
<td>Cycloid</td>
<td>1.474106771</td>
<td>1.825682191</td>
</tr>
</tbody>
</table>
It is an indicator of the poor state of the nature of argumentation at that time in mathematics (especially as it relates to calculus based techniques) that when John gave an argument that the cycloid was indeed the solution to the Brachistochrone problem, James would not accept it, nor could he convince John as what was wrong with the argument, and the dispute led to the eventual alienation of the two brothers. A shame indeed.

Daniel & St. Petersburg

Of the three sons of John Bernoulli, all of whom demonstrated clear mathematical talent: Daniel (1700-1782) is the better known. Born in the Netherlands (while his father was a professor there), but soon afterwards, Daniel moved with his family to Basel (where his father eventually taught). His doctorate degree was in medicine rather than mathematics, and his thesis on the mathematics and physics of breathing is considered one of the earliest in the mathematical aspects of medicine. He followed his older brother Nicholas to St. Petersburg where he stayed 8 years. It was Daniel who attracted Euler (see below) to the Northern capital, but after Nicholas’ death Daniel returned to Basel where he stayed the rest of his life.

As most of the mathematicians of this century, Daniel showed a broad spectrum of mathematical and scientific interests. As mentioned above, he is one of the first authors to write on the mathematics of breathing. He also followed his uncle James’s taste for the relation between mathematics and moral certainty, and utility and what it means. For example, to Daniel it is very clear that the value (or utility) of $1 depends very much on how much money one has—not an earth shaking idea—but he went further, and he speculated that the value could be assigned a function of the wealth, such as the reciprocal of the logarithm of the wealth. He also pursued the idea of error systematically, a topic we will return to when we discuss Laplace. The one topic of Daniel Bernoulli we will delve into a little more depth is the St. Petersburg Paradox.

First, let us consider the following game:

You are to toss a coin. If the first toss is a head, I will pay you $1. If the first toss is not a head, you roll again. If the second toss is a head, I will pay you $2. If you still haven’t rolled a head, you roll again, and if then a head turns up, I will pay you $3, etceteras. How much are you willing to pay to get to play a hand of this game?

Certainly you would be willing to pay $1 since that is the least amount of money you will ever receive from me, but that would be terribly unfair to me since all I have in front of me is a losing proposition. On the other hand, I doubt you would be willing to pay $10 to play since it is very unlikely you will receive that amount or a larger one back. How do we arrive at a fair price is naturally the idea of expectation that we already studied when we looked at Huygens’ contributions. We hence compute your expectation for this game.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>$2</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$3</td>
<td>$\frac{1}{8}$</td>
</tr>
<tr>
<td>$4</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>$5</td>
<td>$\frac{1}{32}$</td>
</tr>
<tr>
<td>$6</td>
<td>$\frac{1}{64}$</td>
</tr>
</tbody>
</table>

...
So your expectation is \[
\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{5}{32} + \frac{6}{64} + \cdots
\]
which if we add a few terms we see it approaches 2. We prove it is indeed 2 by two different techniques. The first is just power series and the geometric series comes to our aid. We start by observing that what we have is nothing else but \[x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \cdots\] evaluated at \(\frac{1}{2}\). But, we know that
\[
\frac{1}{1 - x} = 1 + x + x^2 + x^3 + x^4 + \cdots.
\]
If we take derivatives of both sides we get
\[
\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + 5x^4 + \cdots
\]
and so if we multiply by \(x\), we get
\[
\frac{x}{(1-x)^2} = x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \cdots
\]
If we substitute \(\frac{1}{2}\) for \(x\) in both sides of the equation we get
\[
2 = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{5}{32} + \cdots
\]
Another approach is to let \(M\) (for money) be the expectation. Then as Pascal would have done, \(M = \frac{1}{2} + \frac{1}{2} (M+1)\) since you either win in the first toss (probability \(\frac{1}{2}\)), or you toss again, and then your expectation increases by $1. Solving for \(M\) gives us, \(M = 2\). So if you agree to pay $2 to play, we have an even, or fair, game.

Let us vary the game into the version associated with the St. Petersburg Paradox. Suppose instead that the game is as follows.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1</td>
<td>(\frac{1}{2})</td>
</tr>
<tr>
<td>$2</td>
<td>(\frac{1}{4})</td>
</tr>
<tr>
<td>$4</td>
<td>(\frac{1}{8})</td>
</tr>
<tr>
<td>$8</td>
<td>(\frac{1}{16})</td>
</tr>
<tr>
<td>$16</td>
<td>(\frac{1}{32})</td>
</tr>
<tr>
<td>$32</td>
<td>(\frac{1}{64})</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
</tbody>
</table>

You are to toss a coin. If the first toss is a head, I will pay you $1. If the first toss is not a head, you roll again. If the second toss is a head, I will pay you $2. So far the two games are identical. However, if you still haven't rolled a head, you roll again, and if then a head turns up, I will pay you $4, and in the next roll I would pay you $8, etcetera. How much are you willing to pay to get to play a hand of this game?

The outcomes and probabilities are given by:

And your expectation becomes \[
\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots
\] which is clearly infinite.

If we do á la Pascal, and \(M\) were finite, then \(M = \frac{1}{2} + \frac{1}{2} (2M)\) which gives
\[
\frac{1}{2} = 0,
\]
which is nonsense. Hence, you should be willing to pay any amount of money to play this game. We have nothing paradoxical here. Wherefore the St. Petersburg
paradox? Acquaintances of Daniel ran experiments—what seemed to them a large number of experiments (more than 2,000 tries), and the actual amount of money received during the trial run amounted to approximately $5 on the average, far from the infinity that was expected. Hence the paradox.

However, indeed there is no paradox here. Just a slow game to catch up to the large expectation. The number of games matters tremendously on what the average per hand is going to be. They ran 2,000 and got approximately $5 for an average. But if you go to 1 million hands, you are more likely to get at least $10 for an average, and if you go to 10 million, you will get more than $20. The key, of course, is that there are some very big payoffs in the game, but one may have to wait a long time before one gets such a payoff.

De Moivre & the Bell Shape

Abraham De Moivre (1667-1754) was not particularly fortunate. Not only did he have to flee his native France for England to avoid religious persecution—he was a Huguenot, but once in London, he was never able to secure an academic position and he had to make his living by tutoring young boys—a life that allowed him little time to pursue his research. In the Doctrine of Chances we see many interesting topics involving probability, and he has been called the first great analytic probabilist.

We will discuss three important topics associated with De Moivre.

De Moivre’s Theorem

Although we have seen complex numbers forcing their way into mathematical reality as early as the 1500’s, they did not gain full acceptance until Gauss’ generation, one generation after De Moivre. One of the reasons for acceptance by the latter generation was the geometric connection between the complex numbers and the points of the plane, and the arithmetical operations and geometric transformations which we discussed at the beginning of the notes. De Moivre’s Theorem is certainly a step toward this geometric connection, and perhaps a vital one. Simply stated the theorem claims:

$$\left(\cos(\theta) + i\sin(\theta)\right)^n = \cos(n\theta) + i\sin(n\theta)$$

where $n$ is an arbitrary positive integer and $\theta$ is any angle.

Once the connection between complex multiplication and motion in the plane is understood, De Moivre’s Theorem is geometrically very simple—but, again, we emphasize this connection was not fully understood at De Moivre’s time. We will spend considerable time on the complex numbers in a future chapter, so we will leave this subject for now.

The next topic has someone’s name attached to it—but DeMoivre has done most of the work.
Stirling’s Formula

The factorial had become a fundamental expression in mathematical language (as we will see further in our discussion of Euler below). While in England, De Moivre searched for an approximation to the factorial, an approximation that was improved upon by Stirling, and due to this improvement, it is referred to as Stirling’s Formula:

\[
n! \approx \sqrt{2\pi n} \frac{n^{n+\frac{1}{2}}}{e^n}.
\]

Actually, De Moivre had the formula except for the constant \(\sqrt{2\pi}\) for which he had only given an approximation. In the table, we give a few examples of the two values.

<table>
<thead>
<tr>
<th>(n)</th>
<th>(n!)</th>
<th>(\sqrt{2\pi n} \frac{n^{n+\frac{1}{2}}}{e^n})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.92</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>5.84</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>23.51</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>118.02</td>
</tr>
<tr>
<td>6</td>
<td>720</td>
<td>710.08</td>
</tr>
<tr>
<td>7</td>
<td>5040</td>
<td>4980.40</td>
</tr>
<tr>
<td>8</td>
<td>40320</td>
<td>39902.40</td>
</tr>
<tr>
<td>9</td>
<td>362880</td>
<td>359536.87</td>
</tr>
<tr>
<td>10</td>
<td>3628800</td>
<td>3598695.62</td>
</tr>
<tr>
<td>11</td>
<td>39916800</td>
<td>39615625.05</td>
</tr>
<tr>
<td>12</td>
<td>479001600</td>
<td>475687486.47</td>
</tr>
<tr>
<td>100</td>
<td>9.3326E+157</td>
<td>9.3248E+157</td>
</tr>
</tbody>
</table>

Lastly, we will look at perhaps the most enduring contribution of De Moivre

DeMoivre & the Normal

In the early history of probability, the situation most often encountered was that of performing repeated trials of an experiment, and counting the number of successes among the trials, however success was defined. For example, flipping a coin so many times, and counting the number of heads, or the examples we encountered above when discussing James Bernoulli’s work.

Suppose, for the sake of explicitness, we consider flipping a fair coin 10,000 times, and counting the number of heads. We have seen before that the probability of \(k\) heads among 10,000 tosses of a fair coin is \(\binom{10,000}{k} \frac{1}{2^{10,000}}\), and the probability of any one specific number of successes is very small. To wit, the most likely event, that one of exactly 5,000 heads, has probability less than 0.008, but this computation is not at all trivial even today, thus, so much less so in the eighteenth century. Hence, events have to be collected in groups so as to make quantities meaningful—remember Bernoulli. So we may ask instead, what is the probability of staying within 50 heads of this most likely event, in other words, what is the probability of having between 4,950 and 5,050 heads when we toss a coin 10,000 times. We can readily write an answer, but what numerical estimate we
can attach to it is a different story. The answer is: \[ \sum_{k=4,950}^{5,050} \binom{10,000}{k} \frac{1}{2^{10,000}}. \] But how to arrive at an estimate for such a quantity? And then, if we vary the numbers a little, how sensitive are our estimates to these variations?

De Moivre introduced what is one of the most important distributions into probability, the function \( f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \), which nowadays is so prevalent, it has many names: the normal distribution, the Gaussian density function, the astronomer’s error law, or simply the bell-shaped curve (the same shaped as any row of Pascal’s triangle). And without getting too technical, De Moivre correctly claimed that this unique bell-shaped curve can be used to approximate any of the binomial problems (once appropriately calibrated), and this approximation can be used to give an answer to our query as an integral instead of a sum. In our case, the estimate is: the probability of between 4,950 and 5,050 heads is 0.6826. We will return to the normal when we discuss Laplace in the next chapter.

**The Inimitable Euler**

We end this chapter with a brief description of some of the works of the best mathematician of the eighteenth century, and certainly one of the best of all time: Leonhard Euler (1707-1783). Euler was born in Basel, Switzerland where he studied under John Bernoulli, and befriended his teacher's sons. It was the young Bernoullis that secured a position for Euler in the St. Petersburg Academy in 1727, which he accepted. In 1741, Euler left Russia to go to Berlin and work for Frederick, the Great. Although not the happiest of times, he stayed in Berlin for 25 years. In 1766, he returned to Russia where he stayed until his death. The most prolific of all mathematical writers, Euler was never impaired by his blindness—he lost sight in one eye in 1735, and in both shortly after his return to Russia. Without a doubt, one of the greatest calculators of all time, he could perform impressive feats of calculation in his head. In the latter years, he was aided in his reading and his writing by his daughter. We will look at many contributions of Euler since he is a personal favorite.

**Euler & the 5 Most Important Constants**

We start by mentioning an equation he was very proud of:

\[ e^i\theta + 1 = 0 \]

since it contains the 5 most important numbers. This equation follows trivially from a result he obtained easily by power series:

\[ e^i\theta = \cos \theta + i\sin \theta. \]

The number \( e \) is indeed a number associated with Euler (he named it), and his influence in our notation is indeed massive. He popularized the modern use of \( \pi \). He introduced
the notation $f(x)$ for functions, and he struggled with the concept of function, as did most of his contemporaries. His discussion with D’Alembert about the nature of function in reference to the vibrating string problem helped develop some of the ideas of Fourier.

And since Euler was the first to name the fundamental constant known to us as $e$, it is appropriate we start with a result of his concerning $e$—its irrationality.

**Theorem.** $e$ is irrational.

**Proof.** We know by the Taylor series expansion of the exponential function that 
$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots.$$ Let us suppose, by way of contradiction, that $e$ is not irrational, and hence there exist positive integers $h$ and $k$ such that $e = \frac{h}{k}$. But then, consider the series above as divided into two pieces, a finite sum consisting of those terms whose denominators are less than or equal to $k$, and an infinite sum of all those terms whose denominators exceed $k$, that is
$$\sum_{n=0}^{k} \frac{1}{n!} + \sum_{n=k+1}^{\infty} \frac{1}{n!}.$$ Let us refer to these as $F$ and $R$ respectively. Thus
$$e = F + R,$$ or equivalently,
$$R = e - F.$$

Consider multiplying this equation by $k!$. Clearly, $k!e = (k-1)!h$ is an integer. Also, $k!F = k! + \cdots + \frac{k}{1!} + \frac{k}{2!} + \cdots + \frac{k}{k!}$, being the sum of positive integers, is also a positive integer. Thus we have to conclude that
$$k!F = \frac{k!}{1!} + \frac{k!}{2!} + \cdots + \frac{k!}{k!}$$
is an integer too. Since it is positive, if we can argue that it is less than 1, we will have arrived at a contradiction. But
$$k!R = \frac{k!}{(k+1)!} + \frac{k!}{(k+2)!} + \cdots = \frac{1}{k+1} + \frac{1}{(k+1)(k+2)} + \frac{1}{(k+1)(k+2)(k+3)} + \cdots.$$ However, since $k + 2 > k + 1$,
$$\frac{1}{(k+1)(k+2)} < \frac{1}{(k+1)^2},$$
and, similarly,
$$\frac{1}{(k+1)(k+2)(k+3)} < \frac{1}{(k+1)^3},$$
because $k + 1$ is the smallest of the three factors. In a continuing fashion, we obtain
$$k!R < \frac{1}{k+1} + \frac{1}{(k+1)^2} + \frac{1}{(k+1)^3} + \frac{1}{(k+1)^4} + \cdots =$$
$$= \left(\frac{1}{k+1}\right) \left(1 + \frac{1}{k+1} + \frac{1}{(k+1)^2} + \frac{1}{(k+1)^3} + \cdots\right)$$
and since the second factor is a geometric series, we have
Another interesting issue associated with $e$ was whether $e$ could satisfy some polynomial equation with integer coefficients. This was eventually answered (in the negative) in 1882 by **Lindemann**. And a similar question for $\pi$ needed to be answered in order to put the Quadrature of the circle to rest forever. This was done by **Lindemann** too who proved that both $e$ and $\pi$ were transcendental numbers.

**Euler & the Factorial**

We proceed with an accomplishment in Euler’s youth—the **Generalization of the Factorial**.

By Euler’s time, it had become very clear that the factorial function was for many purposes a fundamental element of mathematical language and expression. Euler set out to generalize it, that is, to find a function defined on as many real numbers as possible so that when a positive integer $n$ is inputted, the output is $n!$. Of course there are many such functions—think graphically, all we are doing is drawing a curve, even a polygonal one, that goes though the points $(n,n!)$ in the plane. Hence we need a rationale for the extension we consider.

We know $n!$ as $1 \times 2 \times 3 \times \cdots \times n$, but we have to have a deeper understanding of the function. One trivial fact is that the factorial is the unique function $g$ defined on the set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ such that

$$g(n+1) = (n+1)g(n) \text{ with } g(1) = 1.$$  

For then, $g(2) = g(1+1) = 2g(1) = 2$, and $g(3) = g(2+1) = 3g(2) = 6$, etcetera. Certainly, one possible rationale then for the extension is to require our new function to satisfy this recursive equation:

$$g(x+1) = (x+1)g(x) \text{ and } g(1) = 1,$$

for any $x$ for which $g$ is defined.

Thus, for example, in typically recursive fashion, if we knew $g\left(\frac{1}{2}\right)$, then we would know $g\left(\frac{3}{2}\right)$ since $g\left(\frac{3}{2}\right) = \frac{3}{2} g\left(\frac{1}{2}\right)$, and so we could also compute $g\left(\frac{5}{2}\right)$, and so on. More importantly, could we compute $g\left(-\frac{1}{2}\right)$? If $-\frac{1}{2}$ is going to be a possible input, then by our equation, $g\left(-\frac{1}{2}\right) = \frac{1}{2} g\left(-\frac{1}{2}\right)$, and so $g\left(-\frac{1}{2}\right) = 2 g\left(-\frac{1}{2}\right)$, and we can use the recursion to compute not only forwards, but also backwards, for now we could also compute $g\left(-\frac{3}{2}\right)$, $g\left(-\frac{5}{2}\right)$, and so on.
Indeed, one of the justifications for $0! = 1$ is exactly obtained this way: $g(0+1) = 1 \times g(0)$, and since $g(1) = 1$, $g(0) = 1$ also. But then, could we also get $g(-1)$? If $-1$ could be an input, then our recursion would force $1 = g(0) = g(-1+1) = 0 \times g(-1)$, and hence this would not work for any value of $g(-1)$. Once $-1$ fails to be an input for the function, by the same reasoning, $-2$, and $-3$, $-4$, $-5$, ... all do too. And so there is no hope of extending the factorial in a reasonable way to the negative integers. However, how about any other values?

When he was 22, Euler came up with the following observation. First, he knew, as we do, that the binomial coefficient is intimately connected with the factorial. Fix a positive integer $n$, and consider, for any $k$, $\binom{k+n}{n} = \binom{k+n}{k} = \frac{(k+n)!}{k!n!}$, then

$$
\frac{1}{\binom{k+n}{n}} = \frac{n!k!}{(n+k)!} = \frac{n!}{(k+1) \times (k+2) \times (k+3) \times \cdots \times (k+n)},
$$

so if we multiply by $k^n$, we have

$$
\frac{k^n}{\binom{k+n}{n}} = n \times \frac{k}{k+1} \times \frac{k}{k+2} \times \frac{k}{k+3} \times \cdots \times \frac{k}{k+n}
$$

and if we let now $k$ grow without bound, since $\frac{k}{k+1}$, $\frac{k}{k+2}$, $\frac{k}{k+3}$, etcetera, all go to 1 as $k$ grows, we obtain

$$
n! = \lim_{k \to \infty} \frac{k^n}{\binom{k+n}{k}} = \lim_{k \to \infty} \frac{k^n \times (k-1) \times \cdots \times 1}{(n+1) \times (n+2) \times \cdots \times (n+k)}.
$$

Ignoring problems with convergence for the time being, can we use this limit to extend our factorial function? Does it satisfy our recursion? So we define

$$
g(x) = \lim_{k \to \infty} \frac{k^x}{\binom{k+x}{k}} = \lim_{k \to \infty} \frac{k^x \times (k-1) \times \cdots \times 1}{(x+1) \times (x+2) \times \cdots \times (x+k)} = \lim_{k \to \infty} a_k(x)
$$

where $a_k(x) = \frac{k^x \times (k-1) \times \cdots \times 1}{(x+1) \times (x+2) \times \cdots \times (x+k)}$. Note that this expression, because of the basic recursion of binomial coefficients that we observed when we studied Newton is quite computable (as we will see below).

Now, $a_k(x+1) = \frac{k^{x+1} \times (k-1) \times \cdots \times 1}{(x+2) \times (x+3) \times \cdots \times (x+1+k)}$

$$
= \frac{a_k(x) \times k \times (x+1)}{x+1+k} = a_k(x)(x+1) \frac{k}{x+1+k},
$$

and since $\frac{k}{x+1+k} \to 1$ as $k \to \infty$, we have the desired recursion: $g(x+1) = (x+1)g(x)$. 
Now, as to the computation. Let us compute \( g\left(\frac{1}{2}\right) \). We will let \( a_k \) stand for \( a_k\left(\frac{1}{2}\right) \). So

\[
a_k = \sqrt{k \times k \times (k-1) \times \cdots \times 1} \left(\frac{3}{2}\right) \times \left(\frac{5}{2}\right) \times \cdots \times \left(\frac{2k+1}{2}\right)
\]

and easily we have

\[
a_{k+1} = \frac{\sqrt{k+1}(k+1)}{\sqrt{k} \left(\frac{2k+3}{2}\right)}.
\]

which can be used effectively as the following table indicates:

<table>
<thead>
<tr>
<th>( n )</th>
<th>( a_n )</th>
<th>( n )</th>
<th>( a_n )</th>
<th>( n )</th>
<th>( a_n )</th>
<th>( n )</th>
<th>( a_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66667</td>
<td>2</td>
<td>0.754247</td>
<td>3</td>
<td>0.791795</td>
<td>4</td>
<td>0.812698</td>
</tr>
<tr>
<td>6</td>
<td>0.835257</td>
<td>7</td>
<td>0.842036</td>
<td>8</td>
<td>0.847222</td>
<td>9</td>
<td>0.851320</td>
</tr>
<tr>
<td>11</td>
<td>0.857380</td>
<td>12</td>
<td>0.859684</td>
<td>13</td>
<td>0.861647</td>
<td>14</td>
<td>0.863340</td>
</tr>
<tr>
<td>16</td>
<td>0.866111</td>
<td>17</td>
<td>0.867259</td>
<td>18</td>
<td>0.868283</td>
<td>19</td>
<td>0.869202</td>
</tr>
<tr>
<td>21</td>
<td>0.870784</td>
<td>22</td>
<td>0.871470</td>
<td>23</td>
<td>0.872097</td>
<td>24</td>
<td>0.872674</td>
</tr>
<tr>
<td>26</td>
<td>0.873696</td>
<td>27</td>
<td>0.874151</td>
<td>28</td>
<td>0.874575</td>
<td>29</td>
<td>0.874969</td>
</tr>
<tr>
<td>31</td>
<td>0.875684</td>
<td>32</td>
<td>0.876008</td>
<td>33</td>
<td>0.876313</td>
<td>34</td>
<td>0.876600</td>
</tr>
<tr>
<td>36</td>
<td>0.877127</td>
<td>37</td>
<td>0.877370</td>
<td>38</td>
<td>0.877600</td>
<td>39</td>
<td>0.877818</td>
</tr>
</tbody>
</table>

And just listing a few of the terms:

<table>
<thead>
<tr>
<th>( n )</th>
<th>( a_n )</th>
<th>( n )</th>
<th>( a_n )</th>
<th>( n )</th>
<th>( a_n )</th>
<th>( n )</th>
<th>( a_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66667</td>
<td>4</td>
<td>0.812698</td>
<td>9</td>
<td>0.851320</td>
<td>16</td>
<td>0.866111</td>
</tr>
<tr>
<td>36</td>
<td>0.877127</td>
<td>49</td>
<td>0.879516</td>
<td>64</td>
<td>0.881076</td>
<td>81</td>
<td>0.882150</td>
</tr>
<tr>
<td>121</td>
<td>0.883492</td>
<td>144</td>
<td>0.883932</td>
<td>169</td>
<td>0.884626</td>
<td>196</td>
<td>0.884986</td>
</tr>
<tr>
<td>256</td>
<td>0.884931</td>
<td>289</td>
<td>0.885079</td>
<td>324</td>
<td>0.885203</td>
<td>361</td>
<td>0.885308</td>
</tr>
<tr>
<td>441</td>
<td>0.885474</td>
<td>484</td>
<td>0.885541</td>
<td>529</td>
<td>0.885599</td>
<td>576</td>
<td>0.885650</td>
</tr>
<tr>
<td>676</td>
<td>0.885736</td>
<td>729</td>
<td>0.885771</td>
<td>784</td>
<td>0.885803</td>
<td>841</td>
<td>0.885832</td>
</tr>
<tr>
<td>961</td>
<td>0.885881</td>
<td>1024</td>
<td>0.885903</td>
<td>1089</td>
<td>0.885922</td>
<td>1156</td>
<td>0.885940</td>
</tr>
<tr>
<td>1296</td>
<td>0.885971</td>
<td>1369</td>
<td>0.885984</td>
<td>1444</td>
<td>0.885997</td>
<td>1521</td>
<td>0.886009</td>
</tr>
</tbody>
</table>

and we are close to the true value of

\[
g\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}.
\]

The actual graph of \( g(x) \) for positive \( x \) is given in the picture:

Years later, Euler would define the **Gamma** function,

\[
\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt.
\]

It suffices to remark that \( \Gamma(x) = g(x+1) \), as we have defined \( g(x) \) in our discussion above.
Euler Invents Graph Theory

It was in 1736 that Euler created graph theory—a subject that has become of some importance in our times. The problem he was trying to solve is known as the **7 Bridges of Königsberg**. In the city of Königsberg there were seven bridges as in the picture:

As a pastime, during their Sunday walk through the city, the townsfolk would try to see if they could take a stroll starting anywhere in the city, and finishing anywhere in the city, but so that they had crossed every bridge exactly once.

Euler realized that to solve this problem, a simplification of the information available was required, and there lies the power of graph theory. A **graph** is nothing else but a collection of **vertices** (or **nodes** or **points**), and a collection of lines joining them. These lines are called **edges** most of the time. Thus in our problem, there are four regions connected by seven edges (the bridges), and thus our problem can be pictured by the following (simpler) graph:

Euler then could easily argue that the desired walk was not possible as follows. Suppose such a walk was possible in a given graph, namely start at some vertex and finish at some vertex, but cross every edge exactly once. Consider any vertex besides the starting and finishing ones (which could be one and the same). Since we are walking every edge exactly once, every time we come into a vertex by an edge, we have to leave that vertex by a different edge, so the number of edges connected to any such vertex is even. Hence we can have our walk through all the edges once only if every vertex, except possibly two, has an even number of edges connected to it. (As it turns out this necessary condition is also sufficient.) In the specific case of the 7 bridges, the vertices have 5, 3, 3 and 3 edges coming out of them, so the walk through every edge exactly once is not possible. Such walks through a graph are referred nowadays as **Eulerian** in honor of our character.

**Euler's Formula**

We extend now graph theoretic reasoning to derive a formula due to Euler (but also perhaps understood by Descartes) concerning any graph that can be drawn in the plane. What do we mean by a **graph that can be drawn in the plane**? We mean that **no edges cross except at a vertex**. For example, if we draw any four points in the plane, and we join any two of them by exactly one edge, so there are exactly $6 = \binom{4}{2}$ edges, then we can do it in the plane as the figure.
shows,

But, we could also have been careless and not drawn it in the plane as this picture shows:

Suppose we draw it differently (but still in the plane) as the next figure shows, then we can do a little counting.

By a **face** we mean an area of the plane surrounded by the edges of a graph (in order for this concept to make sense we have to have drawn the graph in the plane, so the second picture above is not allowed.) We also include the **infinite** region of the plane as one of the faces.

Hence if we count in the first picture, we get 4 regions, and if we count in the third picture, we get 4 faces also. It is perhaps intuitively clear that the number of faces is independent on how we draw the graph as long as we draw it in the plane.

What is perhaps not so clear is a wonderful relation that Euler observed between the number of vertices, edges and faces. If we let $v$ be the number of vertices, so in our previous graph $v = 4$, and if we let $e$ be the number of edges, so $e = 6$, and if we let $f$ be the number of faces, then $v - e + f = 2$.

This is no accident, and indeed it is a theorem as long as the graph can be drawn in the plane and as long as the graph is **connected** (all in one piece). We will use the following notation for graphs: if $x$ and $y$ are vertices, then $x \leftrightarrow y$ will denote the fact that there is an edge joining $x$ and $y$. Thus to be connected means that for any two vertices $x$ and $y$, there is a **path** from $x$ to $y$, namely, a sequence of vertices $x_1, x_2, \ldots, x_n$, each one of them joined to the next, starting at $x$ and ending at $y$:

$$x = x_1 \leftrightarrow x_2 \leftrightarrow \cdots \leftrightarrow x_{n-1} \leftrightarrow x_n = y.$$ 

Certainly the graph associated with the bridges is connected, but the following graph with five vertices is not since it is made up of two not connected pieces:

Before we can prove the aforementioned wonderful theorem, we need some simple definitions and lemmas from graph theory. First, what is a **simple graph**? The meaning is simple, take any (finite) collection of points joined by edges. To be **simple**, it **is required that no two edges join the same two vertices, and no edge joins a vertex with itself.** Note the graph in the bridges problem is not ordinary since there are more than one bridge between different parts of the city, but the disconnected graph above is simple. And so are the graphs with four vertices drawn above. Note also that one can always delete edges from a graph until one has a simple graph, and if the original graph was connected, then so is the new simple graph.
By a cycle in a graph, we mean a path of joined vertices that wraps into itself:

\[ x = x_1 \leftrightarrow x_2 \leftrightarrow \cdots \leftrightarrow x_{n-1} \leftrightarrow x_n = x. \]

The picture illustrates a cycle of length 4. By a nontrivial cycle we mean one length at least 3. The key remark one needs to make about cycles is that if we have a connected graph, and from it we remove an edge in a cycle, then we still have a connected graph, since whatever was accomplished by that edge, can be accomplished the long way around the cycle.

**Lemma 1.** Let \( G \) be a simple graph in which every vertex has at least two edges adjacent to it. Then there is a nontrivial cycle in \( G \).

*Proof.* Take an edge \( x \leftrightarrow y \). Since there are two edges adjacent to \( y \), we must have a new vertex \( z \) such that \( y \leftrightarrow z \). But then there must exist a \( w \) different from \( y \) such that \( z \leftrightarrow w \). If \( w = x \), then we have closed the loop, and we have a cycle of length 3. If, not continue. Eventually, we run out of new points since there are only finitely many, and we must close the loop.

**Lemma 2.** Suppose \( G \) is a graph with \( v \) vertices and \( e \) edges. Suppose that \( G \) is connected. Then \( e \geq v - 1 \).

*Proof.* We can start by assuming \( G \) is simple. By induction on \( v \). Certainly, this is true if \( v = 1 \) or \( v = 2 \). If there is a vertex with only one edge adjacent to it, then remove the vertex and that edge, and we have reduced both numbers by exactly 1, so can apply the induction. If every vertex has at least two edges adjacent to it, then by the previous lemma, we have a cycle. Remove any edge of that cycle, then the new graph is still connected. Keep doing this until we get a vertex with only one edge adjacent to it, and then move to the previous step. Since up until we are only removing edges, we get our inequality.

**Theorem 1.** Let \( G \) be a simple graph. Suppose it has \( v \) vertices and \( e \) edges. Consider the following three conditions:

1. \( G \) is connected.
2. \( G \) has no cycles.
3. \( v = e + 1 \).

Then any two of the conditions implies the third one.

*Proof.* Start by assuming \( 1 \) and \( 2 \). We need to prove \( 3 \). Proceed by induction on \( v \). Obvious if \( v = 2 \). Since there are no cycles, there has to be a vertex adjacent to one edge only. Remove it and the dangling edge. Both conditions \( 1 \) and \( 2 \) are still true. And thus by induction we are done. Assume now \( 1 \) and \( 3 \). Assume there is a cycle. Remove an edge. Then we would still be connected, but now if we let, in the new graph, \( v' = v \) denote the number of vertices, and \( e' = e - 1 \) denote the number of edges, then \( e' = e - 1 = v - 2 < v' - 1 \), contradicting Lemma 2. Finally, assume \( 2 \) and \( 3 \). Then proceed by induction. If \( v = 2 \), then it is clear. It is also obvious if \( v = 3 \). The first interesting case is when \( v = 4 \). If we have then 3 edges and no cycles, two possibilities
arise. If one vertex is adjacent to 3 edges, then we must have: $\bullet$, which is connected. If not, then we are forced into $\bullet$, which is again connected. By Lemma 1, we must have a vertex adjacent to only one edge, and then by removing it and the edge, by induction, we have that the new graph is connected, but then so is the original one.

A graph satisfying the 3 conditions of the Theorem is called a tree.

**Theorem 2.** If $G$ is a connected graph that is drawn in the plane, and $v$ is the number of vertices, $e$, the number of edges, and $f$, the number of faces, then

$$v - e + f = 2.$$

*Proof.* First we argue that we can assume that the graph is simple. For if not, any two edges between the same two vertices gives a face, so if we remove one of those double edges, $f \mapsto f - 1$, but also $e \mapsto e - 1$, and so our relation remains unchanged. We now proceed by induction on the number of faces. If there is only one face, then there are no cycles, and then by Theorem 1, $v - e = 1$, and so since $f = 1$, we have the theorem. Otherwise take any face, and remove an edge from it. What has occurred now is that we still have a connected graph since we removed an edge from a cycle, but we also have changed $f \mapsto f - 1$ and $e \mapsto e - 1$. And thus by induction, we are done.

For another example of verification of the theorem, consider the following graph where $v=8$, $e=12$, and $f=6$. If the numbers are reminiscent of something else, indeed the cube, or hexahedron, has the same number of vertices, edges and faces. Actually what Euler was interested in was convex polyhedra in space. But we can easily go from convex polyhedra to graphs one can draw on a sphere.

This is accomplished by thinking of the cube in the example above as being made out of rubber and blowing it into a ball, as in the picture below. Then we can think of the plane as being wrapped around the ball, in very much the same fashion as we make maps out of our spherical Earth. When we do that we get the graph drawn in the plane above.

There are distortions, for sure, but the count of vertices, edges and faces stays the same,
and hence any convex polyhedron satisfies

\[ v - e + f = 2. \]

This is often referred to as Euler’s formula.

Once we have Euler’s formula, it is very easy to classify the regular or platonic solids. We will prove the result in the exercises. Of course their shapes are well known.

---

**Euler & the Officers**

We take a short combinatorial detour of Euler’s interests: the 36 Officer Problem

Consider the following arrangement of 16 distinct army officers, one of each of 4 different ranks: 1, 2, 3, & 4 and of 4 different regiments: ⊙, →, ⊐, & ⊓.

<table>
<thead>
<tr>
<th>1⊙</th>
<th>2→</th>
<th>3⊐</th>
<th>4 mộ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2⊐</td>
<td>1⊙</td>
<td>4⊙</td>
<td>3→</td>
</tr>
<tr>
<td>3⊓</td>
<td>4⊐</td>
<td>1→</td>
<td>2⊙</td>
</tr>
<tr>
<td>4→</td>
<td>3⊙</td>
<td>2⊓</td>
<td>1⊐</td>
</tr>
</tbody>
</table>

We easily see that each row has every rank represented and every regiment represented, and the same is true of every column. Can we do the same with 36 officers, one of each of six ranks and one of each of six regiments? Euler played with this, and since he could devise such an arrangement for any number \( n \) of officers except for \( n=6, 10, 14, 18, 22, \ldots \), he concluded this could not be done. Eventually he was proven right about \( n = 6 \), but wrong about all others!
Euler & Number Theory

Euler had a tremendous effect on Number Theory. With the possible exception of Fermat, Euler is utmost in his influence in this field. Indeed he prove Fermat’s last theorem for the case $n = 3$. The area in which we will look at him is in The generalization of Fermat’s Little Theorem and the Euler $\phi$-function.

One of Fermat’s useful results is his famous Little Theorem:

\[
\text{if } p \text{ is a prime, then } n^p - n \text{ is a multiple of } p.
\]

Indeed, Euler was the first one to prove this, and his proof by induction is very nice indeed.

Euler set out to generalize this theorem, which is not in general true if $p$ is not a prime—$5^6 - 5$ is not a multiple of 6. In the long process of trying to find the correct statement that was true, Euler encountered a function that still carries his name today, the Euler $\phi$-function. The function $\phi$ counts the number of integers relatively prime to a given integer. More formally, $\phi(n)$ is the number of integers $\leq n$ which are relatively prime to $n$. For example, $\phi(2) = 1$, since 1 is the only integer in consideration, $\phi(3) = 2$ since both 1 and 2 are relatively prime to 3, $\phi(4) = 2$ since 1 and 3 are relatively prime to 4, $\phi(5) = 4$ since all 1,2,3, and 4 are relatively prime to 5, but $\phi(6) = 2$ since 1 and 5 are the only ones to be counted. Furthermore,

\[
\begin{array}{c|c|c}
 n & \text{integers } \leq n \text{ which are relatively prime to } n & \phi(n) \\
1 & 1 & 1 \\
3 & 1,2 & 2 \\
5 & 1,2,3,4 & 4 \\
7 & 1,2,3,4,5,6 & 6 \\
9 & 1,2,4,5,7,8 & 6 \\
11 & 1,2,3,4,5,6,7,8,9,10 & 10 \\
13 & 1,2,3,4,5,6,7,8,9,10,11,12 & 12 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
 n & \text{integers } \leq n \text{ which are relatively prime to } n & \phi(n) \\
2 & 1 & 1 \\
4 & 1,3 & 2 \\
6 & 1,5 & 2 \\
8 & 1,3,5,7 & 4 \\
10 & 1,3,7,9 & 4 \\
12 & 1,5,7,11 & 4 \\
14 & 1,3,5,9,11,13 & 6 \\
\end{array}
\]

Easily, for any prime $p$, $\phi(p) = p - 1$, and Fermat's Little theorem can be restated in the form:

\[
\text{if } n \text{ is relatively prime to the prime } p, \text{ then } n^{p-1} - 1 \text{ is a multiple of } p.
\]

Euler's theorem is then a generalization of this statement:

**Euler's Theorem.** Let $n$ be relatively prime to the positive integer $m$. Then $n^{\phi(m)} - 1$ is a multiple of $m$.

For example, since 5 is relatively prime to 6, and $\phi(6) = 2$, then $5^2 - 1$ is a multiple of 6 (which is true), or since 7 is relatively prime to 12, and $\phi(12) = 4$, then $7^4 - 1 = 2400$ is a multiple of 12.
Euler & the Zeta Function

Now we arrive at a famous success by our great mathematician: The Zeta Function. Nowadays, we define the zeta function of a complex variable by $\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$ whenever the series converges. Thus we know that $\zeta(1)$ is not defined (the harmonic series diverges), and Euler computed $\zeta(2), \zeta(4), \ldots, \zeta(26)$. The zeta function evaluated at the odd positive integers is very mysterious, only in the 1980’s, $\zeta(3)$ was shown to be irrational. We still do not know whether it is transcendental or not.

Euler was the first one to be able to get the sum of the reciprocals of the squares, a result that had eluded the Bernoullis, or equivalently to compute $\zeta(2)$. We give a loose argument based on Euler, who used simple tools such as Viéte’s relations in order to arrive at the result. But before we get into the proof, we will need to rethink Viéte’s relations to fit closer to the eighteenth century. We will be spending more time on these in the next chapter when we study Lagrange.

We are accustomed to writing polynomials in the form where the leading term is the term with the highest degree, and when we factor we tend to write $(x - \alpha)$ as one of the factors whenever $\alpha$ is one of the roots. For example, the polynomial with roots 1, 2 and 3 would normally be written in the form $x^3 - 6x^2 + 11x - 6 = (x - 1)(x - 2)(x - 3)$. We tend to prefer monic polynomials, that is those with leading coefficient 1, so, for example, although $2x^3 - 12x^2 + 22x - 12$ also has roots 1, 2, 3, we would prefer to think of the latter as twice the previous monic polynomial.

But there is another natural way to write polynomials. Namely, starting with the constant term, and if as usual, we would prefer to make this leading coefficient 1, we could rather have $1 - \frac{11}{6} x + x^2 - \frac{1}{6} x^3$ instead of $x^3 - 6x^2 + 11x - 6$. Note, first, that this approach is more conducive to thinking of power series—which to Euler were nothing but long polynomials. Second, when we think of factoring such a polynomial, we don't need the roots, rather we need the reciprocals of the roots. To see this we exemplify with a general cubic.

Suppose $x^3 - ax^2 + bx - c$ is a general cubic, and suppose $\alpha, \beta, \gamma$ are the roots, then we know

\begin{align*}
a &= \alpha + \beta + \gamma, \\
b &= \alpha \beta + \alpha \gamma + \beta \gamma \text{ and} \\
c &= \alpha \beta \gamma
\end{align*}

(note we adjusted the signs ahead of time) since $(x - \alpha)(x - \beta)(x - \gamma) = x^3 - ax^2 + bx - c$. If we are to write this polynomial starting with the constant term first and dividing by $c$, we get
\[
1 - b \frac{x}{c} + \frac{a}{c} x^2 - \frac{1}{c} x^3 = -\frac{1}{\alpha \beta \gamma} (x - \alpha)(x - \beta)(x - \gamma)
= (-1) \left( \frac{1}{\alpha} \right) (x - \alpha)(-1) \left( \frac{1}{\beta} \right) (x - \beta)(-1) \left( \frac{1}{\gamma} \right) (x - \gamma) = \left( 1 - \frac{x}{\alpha} \right) \left( 1 - \frac{x}{\beta} \right) \left( 1 - \frac{x}{\gamma} \right).
\]

This exemplifies the following general principle:

**A Generalized (\& Slightly Different) Version of Viète’s Relations:** Let

\[1 - ax + bx^2 - cx^3 + dx^4 - ex^5 + \cdots\] have for its roots \(\alpha, \beta, \gamma, \delta, \ldots\). Then

\[
a = \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma} + \cdots,
\]

\[
b = \frac{1}{\alpha \beta} + \frac{1}{\alpha \gamma} + \frac{1}{\beta \gamma} + \frac{1}{\alpha \delta} + \frac{1}{\beta \delta} + \cdots,
\]

\[
c = \frac{1}{\alpha \beta \gamma} + \frac{1}{\alpha \beta \delta} + \cdots,\text{ etcetera.}
\]

The reason for this stems from the factorization of

\[1 - ax + bx^2 - cx^3 + dx^4 - \cdots = \left( 1 - \frac{x}{\alpha} \right) \left( 1 - \frac{x}{\beta} \right) \left( 1 - \frac{x}{\gamma} \right) \ldots\]

Once we accept this, we can easily get that the **sum of the reciprocals of the squares** is \(\frac{\pi^2}{6}\). The reasoning is as follows:

Consider the power series expansion of

\[
\frac{\sin(x)}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots.
\]

The zeros of this power series are \(\pm \pi, \pm 2\pi, \pm 3\pi, \pm 4\pi, \text{ etcetera.}\) Now we know that the sum of the reciprocals of the roots equals the coefficient of \(x\)—which is 0. This is not surprising since every root comes accompanied by its negative, so they easily cancel each other. But in order to expedite things, we should look at the factorization we take as given:

\[
1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots = \left( 1 - \frac{x}{\pi} \right) \left( 1 + \frac{x}{\pi} \right) \left( 1 - \frac{x}{2\pi} \right) \left( 1 + \frac{x}{2\pi} \right) \ldots.
\]

If we regroup our factors we will have

\[
1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots = \left( 1 - \frac{x^2}{\pi^2} \right) \left( 1 - \frac{x^2}{4\pi^2} \right) \left( 1 - \frac{x^2}{9\pi^2} \right) \cdots. \quad (\star)
\]

And if we expand the right-hand side, and collect the coefficient of \(x^2\), we immediately see that

\[
\frac{1}{\pi^2} + \frac{1}{2^2 \pi^2} + \frac{1}{3^2 \pi^2} + \frac{1}{4^2 \pi^2} + \cdots = \frac{1}{3!} = \frac{1}{6}.
\]

And multiplying by \(\pi^2\), we get our result:
\[ \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}. \]

We can push this simple algebra to compute other sums of reciprocal powers, just as Euler did. For example, the coefficient of \( x^4 \) on the left-hand side of (\( \star \)) is \( \frac{1}{120} = \frac{1}{5!} \), but the coefficient on the right-hand side is computed by multiplying different coefficients of \( x^2 \), and so it equals \( \sum_{m=1}^{\infty} \frac{1}{m^2} \). Equating the two sides, we have that

\[ \sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^4}{120}. \]

But we have already established that \( \sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^2}{6}. \) We also need that

\[ (A + B + \Gamma + \Delta + E + \cdots)^2 = A^2 + B^2 + \Gamma^2 + \Delta^2 + E^2 + \cdots \]

\[ + 2(AB + A\Gamma + A\Delta + A\Gamma + A\Delta + \cdots + B\Gamma + B\Delta + B\Gamma + B\Delta + \cdots) \]

—or in words, the square of a sum is the sum of the squares plus twice the sum of the products of two terms at a time.

If we square the equation: \( \sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^2}{6} \), we have \( \left( \sum_{m=1}^{\infty} \frac{1}{m^2} \right)^2 = \left( \frac{\pi^2}{6} \right)^2 \), and so \( \sum_{m=1}^{\infty} \frac{1}{m^4} \)

\[ + 2 \sum_{m<n} \frac{1}{m^2 n^2} = \frac{\pi^4}{36}, \]

from which we obtain

\[ \sum_{m=1}^{\infty} \frac{1}{m^4} = \frac{\pi^4}{36} - 2 \sum_{m<n} \frac{1}{m^2 n^2} = \frac{\pi^4}{36} - \frac{2\pi^4}{120} = \frac{(10-6)\pi^4}{360} = \frac{\pi^4}{90}. \]

Euler continued in this way and obtained \( \sum_{m=1}^{\infty} \frac{1}{m^{2k}} \) for \( k = 1, \ldots, 13 \), all of them being of the form \( \frac{m\pi^{2k}}{n} \) where \( m \) and \( n \) are integers.

The mathematician most closely associated with the zeta function is Riemann. Some of the deeper properties of the function were discovered by him. One of the most remarkable is the following:

\[ 2^{1-z} \Gamma(z) \zeta(z) \cos \left( \frac{z\pi}{2} \right) = \pi^z \zeta(1-z) \]

where \( \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \) is the ubiquitous gamma function of Euler.

Perhaps the most famous open problem in mathematics at present, and with many more consequences than Fermat’s last theorem is known as the Riemann Hypothesis and concerns the zeroes of the zeta function. It can be proven that if \( z = a + bi \) is a complex such that \( \zeta(z) = 0 \) then either \( a \) is a negative even integer and \( b = 0 \), or \( 0 \leq a \leq 1 \). Riemann conjectured that indeed
if \( \zeta(z) = 0 \) and \( z \) is not a negative integer, then \( a = \frac{1}{2} \).

Hardy (1914) proved that indeed infinitely many zeroes of the zeta function lie on the vertical line \( a = \frac{1}{2} \), and indeed more than a billion and half zeroes have been found and they all lie on the line.

The Riemann hypothesis has strong consequences as to the distribution of primes. That the latter is connected to the zeta function is not surprising since Euler proved that:

\[
\prod_{p \text{ prime}} \left(1 - \frac{1}{p^z}\right) = \frac{1}{\zeta(z)}.
\]

**Euler & the sum of the Reciprocals of the Primes**

On a similar theme, Euler was the first one to prove that the sum of the reciprocals of the prime numbers is infinite—the series \( \sum_{p \text{ prime}} \frac{1}{p} \) diverges. The key to Euler’s proof of this fact is the following lemma:

**Lemma.** Let \( a \) be a real number with \( 0 \leq a \leq \frac{1}{2} \), then \( \frac{1}{1-a} \leq e^{a+a^2} \).

That this is the case can easily be seen from the graph of the two functions, but a more formal proof can easily be given by taking the natural logarithm of both sides.

So we need to prove \( \ln \left( \frac{1}{1-a} \right) \leq a + a^2 \) for any \( a \) that satisfies \( 0 \leq a \leq \frac{1}{2} \). But the series expansion of

\[
\ln \left( \frac{1}{1-a} \right) = a + \frac{a^2}{2} + \frac{a^3}{3} + \frac{a^4}{4} + \cdots
\]

\[= a + a^2 \left( \frac{1 + a}{2} + \frac{a^2}{3} + \frac{a^3}{4} + \cdots \right) \leq a + \frac{a^2}{2} \left( 1 + a^2 + a^3 + \cdots \right) = a + \frac{a^2}{2} \left( \frac{1}{1-a} \right)
\]
and since \( 0 \leq a \leq \frac{1}{2} \), \( \frac{1}{1-a} \leq 2 \), and we have our lemma.

Now we proceed to prove that the sum of the reciprocals of the primes is infinite.

Suppose otherwise, \( \sum_{p \text{ prime}} \frac{1}{p} = \alpha \), for some positive number \( \alpha \). But we also know that
For any prime $p$, let $a = \frac{1}{p}$ in the previous lemma. Then the sum of all the reciprocals of all the powers of $p$,

$$1 + \frac{1}{p^2} + \frac{1}{p^3} + \cdots = \frac{1}{1 - p} \leq e^{p^{1/p}} = e^a e^{\frac{1}{p}}.$$

We need to see what happens when we multiply two of these series for primes $p$ and $q$, for example:

$$\left(1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots\right) \left(1 + \frac{1}{q} + \frac{1}{q^2} + \frac{1}{q^3} + \cdots\right) = 1 + \frac{1}{p} + \frac{1}{q} + \frac{1}{pq} + \frac{1}{p^2} + \frac{1}{q^2} + \cdots$$

and it should be clear that what we get is the sum of all the reciprocals of all positive integers that are divisible by only $p$'s and $q$'s.

Similarly, if we multiplied three of the series for the primes $p$, $q$ and $r$, for example, then we would get the sum of all the reciprocals of all positive integers that are divisible by only $p$'s, $q$'s and $r$'s.

Continuing in this fashion, we get the sum of all reciprocals of all positive integers factorable into primes (in other words all positive integers) if we keep multiplying, thus:

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p \text{ prime}} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) = \prod_{p \text{ prime}} \left(\frac{1}{1 - p}\right) \leq \prod_{p \text{ prime}} e^{\frac{1}{p}} \leq \left(\prod_{p \text{ prime}} e^p\right) \left(\prod_{p \text{ prime}} e^{\frac{1}{p^2}}\right) = e^{\sum_{p \text{ prime}} \frac{1}{p}} e^{\sum_{p \text{ prime}} \frac{1}{p^2}} = e^{\alpha} e^{\frac{\pi^2}{6}} < \infty,$$

which is nonsense, since we know that the harmonic series diverges.