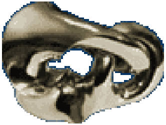
MATH3500
The 6th Millennium Prize Problem

RH



Some numbers have the special property that they cannot be expressed as the product of two smaller numbers, e.g., 2, 3, 5, 7, etc. Such numbers are called prime numbers, and they play an important role, both in pure mathematics and its applications. The distribution of such prime numbers among all natural numbers does not follow any regular pattern, however the German mathematician G.F.B. Riemann (1826–1866) observed that the frequency of prime numbers is very closely related to the behavior of an elaborate function

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots$$

called the Riemann zeta function.

The Riemann hypothesis asserts that all interesting solutions of the equation

$$\zeta(s) = 0$$

lie on a certain vertical straight line. This has been checked for the first $1,500,000,000$ solutions. A proof that it is true for every interesting solution would shed light on many of the mysteries surrounding the distribution of prime numbers.

The primes and their distribution

Definition. A prime number is a natural number greater than 1 with exactly 2 natural numbers as its divisors: 1 and the number itself.

The first 91 prime numbers are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71,$$
$$73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149,$$
$$151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227$$
$$229, 233, 239, 241, 251, 257, 263, 269, 271, 277, 281, 283, 293, 307,$$
$$311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389,$$
$$397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467$$

Around 250 BC Euclid of Alexandria showed that no largest prime exist.

There is an infinity of primes

There are countless proofs of this fact. Below we give a proof attributed to Ernst Kummer, famous for his work on Fermat's Last Theorem.

Proof: As in Euclid's well-known proof, Kummer's proof is a proof by contradiction.

Assume the set $\mathbb{P} = \{2, 3, 5, 7, 11, \dots\}$ of all primes is finite, i.e., there exists a positive integer k such that $\mathbb{P} = \{p_1, \dots, p_k\}$. From our previous table we already know that $k \geq 91$.

Let $N = p_1 p_2 \cdots p_k$. Clearly $p_k < N - 1$. Hence $N - 1$ is composite, and there exists a $p \in \mathbb{P}$ such that $p|N - 1$ (read: p divides $N - 1$). But also $p|N$. Since, for $N > M$,

$$a|N \quad \text{and} \quad a|M \quad \text{implies} \quad a|N - M,$$

this implies that $p|1$. Since $p > 1$ we have arrived at a contradiction. □

Later we will give another proof due to Leonhard Euler.

One of the most important functions in the study of the primes is the prime-counting function $\pi(x) := \{p \in \mathbb{P} : \ p < x\}$, which gives the number of primes less than x.

For example,

$$\begin{aligned}
\pi(10^3) &= 168 \\
\pi(10^6) &= 78,498 \\
\pi(10^9) &= 50,847,534 \\
\pi(10^{12}) &= 37,607,912,018 \\
\pi(10^{15}) &= 29,844,570,422,669 \\
\pi(10^{18}) &= 24,739,954,287,740,860
\end{aligned}$$

Around 1800 Adrien–Marie Legendre and Carl Friedrich Gauss used tables of the primes to conjecture what now goes by the name of the prime number theorem (PNT):

$$\pi(x) \sim \frac{x}{\log x} \qquad \text{as } x \to \infty$$

Neither Legendre nor Gauss formulated the PNT as above, and the many versions of the theorem is due to the transitivity of the $\sim$ symbol:

If $f(x) \sim g(x)$ and $g(x) \sim h(x)$ then $f(x) \sim h(x)$.

x	$\pi(x)$	$x/\log x$	ratio
10^3	168	144.8	1.1605
10^6	78,498	72,382.4	1.0845
10^9	50,847,534	48,254,942.4	1.0537
10^{12}	37,607,912,018	36,191,206,825.3	1.0391
10^{15}	29,844,570,422,669	28,952,965,460,216.8	1.0308
10^{18}	24,739,954,287,740,860	24,127,471,216,847,323.8	1.0254

Gauss in fact used the logarithmic integral

$$\mathrm{Li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t} \sim \frac{x}{\log x}$$

to formulate the PNT as:

$$\pi(x) \sim \mathrm{Li}(x) \qquad \text{as } x \to \infty$$

x	$\pi(x)$	$\mathrm{Li}(x)$	ratio
10^3	168	176.6	1.051
10^6	78,498	78,626.5	1.0016
10^9	50,847,534	50,849,233.9	1.000033
10^{12}	37,607,912,018	37,607,950,279.8	1.0000010
10^{15}	29,844,570,422,669	29,844,571,475,286.5	1.000000035
10^{18}	24,739,954,287,740,860	24,739,954,309,690,414.0	1.0000000089

It was long believed that $\mathrm{Li}(x) > p(x)$ for all x, but in 1914 John Edensor Littlewood proved that $\mathrm{Li}(x) - p(x)$ changes sign infinitely often. We now know that the first sign-change *must* occur for

$$x < 1.39822 \times 10^{316}.$$

This is a big improvement on an earlier bound of

$$x < 10^{10^{10^{34}}}.$$

Expressing the PNT using the logarithmic integral does more than to "just" give a better approximation of $\pi(x)$.

It gives rise to a probabilistic way of viewing the primes, and (very) loosely speaking:

> The probability of the number x being a prime is $\dfrac{1}{\log x}$

This should by no means be taken literally; the probability of

$$1909834783456284562084562084562038 45$$

being a prime is 0.

As a justification of the above probabilistic viewpoint, assume that a natural number n is prim (whatever that means) with probability $\frac{1}{\log n}$ (i.e., you throw a die with $\log n \neq \mathbb{N}$ sides and if 1 comes up you term n a prim and if one of the other $-1 + \log n$ numbers comes up you term it composit).

Then the expected number of prims $\Pi(x)$ less than or equal to x is

$$\Pi(x) = \sum_{n=2}^{x} \frac{1}{\log n}.$$

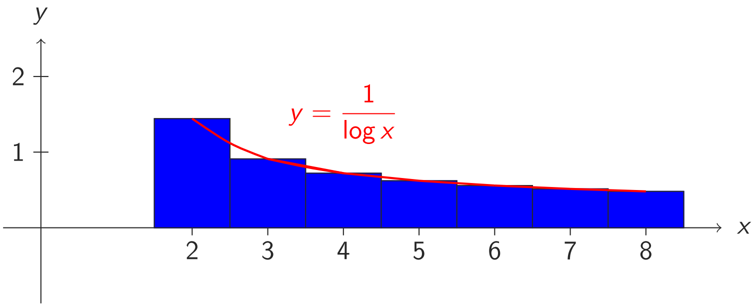
x	$\pi(x)$	$\Pi(x)$	ratio
10^3	168	176.4	1.056183335
10^6	78,498	78,627.3	1.001647712
10^9	50,847,534	50,849,234.7	1.000033433

Alternatively, you can directly assume a "continuum model" (i.e., every $x \in \mathbb{R}$ with $x \geq 2$ is either prim or composit with probability $\dfrac{1}{\log x}$ and $1 - \dfrac{1}{\log x}$. Then $\pi(x)$, the expected number of prims less than x, is

$$\pi(x) = \int_2^x \frac{\mathrm{d}t}{\log t} = \mathrm{Li}(x).$$

Note in particular that

$$\Pi(x) = \sum_{n=2}^{x} \frac{1}{\log n} \sim \int_2^x \frac{\mathrm{d}t}{\log t} = \mathrm{Li}(x).$$



"Proof" that $\displaystyle\sum_{n=2}^{x} \frac{1}{\log n} \sim \int_2^x \frac{\mathrm{d}t}{\log t}$.

Since

$$\mathrm{Li}(x \log x) \sim x$$

the PNT is also equivalent to the statement that

$$p_n \sim n \log n.$$

In other words, for large n, one expects the nth prime number p_n to occur roughly around $n \log n$.

n	p_n	$n \log n$	ratio
10^3	7,919	6907.8	1.146392667
10^6	15,485,863	13,815,510.6	1.120904141
10^9	22,801,763,489	20,723,265,836,946.4	1.100297785

Without recourse to the PNT Euler already established results for the density of primes which, as a by-product, implied another proof of the infinity of primes.

First let us recall the Harmonic series

$$\zeta(1) := \sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$$

This series sums the reciprocals of all of the positive integers.

Amazingly, Bishop of Lisieux Nicole Oresme ($\pm$ 1323–1382) already proved:

The harmonic series diverges

Proof: Let

$$\zeta_N(1) := \sum_{n=1}^{2^N} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^N}$$

Then

$$\begin{aligned}
\zeta_N(1) &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{1+2^{N-1}} + \cdots + \frac{1}{2^N}\right) \\
&> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^N} + \cdots + \frac{1}{2^N}\right) \\
&= 1 + \underbrace{\frac{1}{2} + \cdots + \frac{1}{2}}_{N \text{ times}} \\
&= 1 + \frac{1}{2} N.
\end{aligned}$$

Hence "$\lim_{N \to \infty} \zeta_N(1) = \infty$". □

Now what happens if we sum the reciprocals of the squares, i.e., what (if any) is the value of

$$\zeta(2) := \sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \cdots ?$$

This was first answered by Euler:

$$\zeta(2) = \frac{\pi^2}{6}$$

Proof: Note that

$$\int_{[0,1]^2} (xy)^{2n} \mathrm{d}x\mathrm{d}y = \left(\int_0^1 x^{2n} \mathrm{d}x \right)^2 = \frac{1}{(2n+1)^2}.$$

Hence

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} &= \sum_{n=0}^{\infty} \int_{[0,1]^2} (xy)^{2n} \mathrm{d}x\mathrm{d}y \\ &= \int_{[0,1]^2} \sum_{n=0}^{\infty} (xy)^{2n} \mathrm{d}x\mathrm{d}y \\ &= \int_{[0,1]^2} \frac{1}{1-(xy)^2} \mathrm{d}x\mathrm{d}y. \end{aligned}$$

To evaluate the integral, the trick is to change variables from x, y to u, v as follows

$$x = \frac{\sin u}{\cos v} \quad \text{and} \quad y = \frac{\sin v}{\cos u}.$$

The Jacobian of this variable change is

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \det \begin{pmatrix} \frac{\cos u}{\cos v} & \frac{\sin u \sin v}{\cos^2 v} \\ \frac{\sin u \sin v}{\cos^2 u} & \frac{\cos v}{\cos u} \end{pmatrix} = 1 - \tan^2 u \tan^2 v.$$

Since, miraculously,

$$\frac{1}{1-(xy)^2} = \frac{1}{1-\tan^2 u \tan^2 v}$$

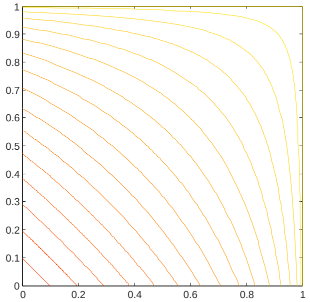
we find

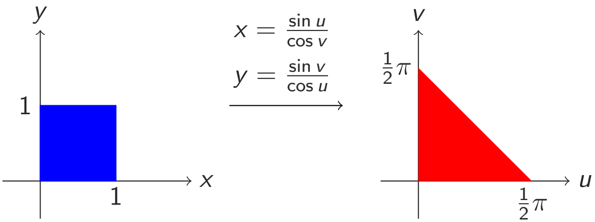
$$\int_{[0,1]^2} \frac{1}{1-(xy)^2} \, dxdy = \int_{??} 1 du dv = \text{Area of ??}.$$

The curves

$$(x(u,v),y(u,v)) = \left(\frac{\sin u}{\cos v}, \frac{\sin v}{\cos u}\right), \qquad u+v=A,\ u,v\geq 0,$$

for $A = \frac{1}{32}\pi, \frac{2}{32}\pi, \ldots, \frac{15}{32}\pi$:





Obviously, the area of the red triangle is $\frac{\pi^2}{8}$, so that

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \int_{[0,1]^2} \frac{1}{1-(xy)^2}\,\mathrm{d}x\mathrm{d}y = \frac{\pi^2}{8}.$$

This almost completes the proof. Indeed,

$$\begin{aligned}\zeta(2) &= \sum_{\substack{n=1 \\ n \text{ even}}} \frac{1}{n^2} + \sum_{\substack{n=1 \\ n \text{ odd}}} \frac{1}{n^2} \\ &= \sum_{n=1}^{\infty} \frac{1}{(2n)^2} + \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \\ &= \frac{1}{4}\zeta(2) + \frac{\pi^2}{8}\end{aligned}$$

which implies that

$$\zeta(2) = \frac{\pi^2}{6}.$$

□

What do $\zeta(1)$ and $\zeta(2)$ have to do with the density of the primes?

Obviously the primes are "not as dense" as the natural numbers.
But are they "denser" or "less dense" than the squares of the natural numbers?

We have just seen that the squares are significantly less dense than the naturals, since $\zeta(1) = \infty$ and $\zeta(2) = \frac{\pi^2}{6}$.

Euler wondered if the sum of the reciprocals of the primes has a finite sum, i.e., if

$$\sum_{p \in \mathbb{P}} \frac{1}{p}$$

converges.

To answer this question he proved the now famous Euler product formula, which connects the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

to the prime numbers:

$$\zeta(s) = \prod_{p\in\mathbb{P}} \frac{1}{1-p^{-s}} \quad \text{for} \quad \mathrm{Re}(s) > 1$$

Proof: Recall the geometric series

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \text{for} \quad |x| < 1.$$

Now truncate Euler's product formula and define

$$\zeta(s;x) := \prod_{p \in \mathbb{P}_x} \frac{1}{1-p^{-s}}$$

where $\mathbb{P}_x$ is the set of primes less than or equal to x:

$$\mathbb{P}_x := \{p \in \mathbb{P} : \; p \leq x\}.$$

For each $p \in \mathbb{P}_x$ and $\text{Re}(s) > 1$ we have $0 < |1/p^s| < 1$ so that we can apply the geometric series to each term in the product:

$$\zeta(s;x) = \prod_{p \in \mathbb{P}_x} \sum_{n=0}^{\infty} p^{-ns}.$$

Let m be the cardinality of $\mathbb{P}_x$, i.e., $\mathbb{P}_x = \{p_1, p_2, \ldots, p_m\}$. Then

$$\begin{aligned}\zeta(s;x) &= \prod_{i=1}^{m}\sum_{n=1}^{\infty} p_i^{-ns} \\ &= \left(1 + \frac{1}{p_1^s} + \frac{1}{p_1^{2s}} + \cdots\right) \times \left(1 + \frac{1}{p_2^s} + \frac{1}{p_2^{2s}} + \cdots\right) \\ &\qquad \times \cdots \times \left(1 + \frac{1}{p_m^s} + \frac{1}{p_m^{2s}} + \cdots\right).\end{aligned}$$

Fully expanding the right-hand side gives

$$\begin{aligned}\zeta(s;x) = 1 + \sum_{i=1}^{m} \frac{1}{p_i^s} + \sum_{1\le i\le j\le m} \frac{1}{(p_i p_j)^s} \\ + \sum_{1\le i\le j\le k\le m} \frac{1}{(p_i p_j p_k)^s} + \cdots\end{aligned}$$

Next recall the Fundamental Theorem of Arithmetic: each natural number n greater than 1 has a unique prime factorisation;

$$n = p_1 p_2 \cdots p_k,$$

where $p_1, \ldots, p_k \in \mathbb{P}$ and $p_1 \leq p_2 \leq \cdots \leq p_k$.

For example,

$$\begin{aligned}
10 &= 2 \times 5 \\
11 &= 11 \\
12 &= 2 \times 2 \times 3 \\
13 &= 13 \\
14 &= 2 \times 7 \\
15 &= 3 \times 5 \\
16 &= 2 \times 2 \times 2 \times 2 \\
17 &= 17 \\
18 &= 2 \times 3 \times 3 \\
19 &= 19 \\
20 &= 2 \times 2 \times 5.
\end{aligned}$$

Hence

$$\sum_{i=1}^{m} \frac{1}{p_i^s} = \sum_{\substack{n=2 \\ n \text{ has exactly 1 prime factor} \\ n \text{ has no prime factors exceeding } p_m}}^{\infty} \frac{1}{n^s}.$$

Similarly,

$$\sum_{1 \le i \le j \le m} \frac{1}{(p_i p_j)^s} = \sum_{\substack{n=2 \\ n \text{ has exactly 2 prime factors} \\ n \text{ has no prime factors exceeding } p_m}}^{\infty} \frac{1}{n^s}$$

and

$$\sum_{1 \le i \le j \le k \le m} \frac{1}{(p_i p_j p_k)^s} = \sum_{\substack{n=2 \\ n \text{ has exactly 3 prime factors} \\ n \text{ has no prime factors exceeding } p_m}}^{\infty} \frac{1}{n^s}.$$

We may thus conclude that

$$\zeta(s;x) = 1 + \sum_{i=1}^{m} \frac{1}{p_i^s} + \sum_{1\le i\le j\le m} \frac{1}{(p_i p_j)^s} + \sum_{1\le i\le j\le k\le m} \frac{1}{(p_i p_j p_k)^s} + \cdots$$

$$= 1 + \sum_{\substack{n=2 \\ n \text{ has no prime factors exceeding } p_m}}^{\infty} \frac{1}{n^s}$$

$$= 1 + \sum_{\substack{n=2 \\ n \text{ has no prime factors exceeding } x}}^{\infty} \frac{1}{n^s}.$$

To summarise:

$$\prod_{p\in\mathbb{P}_x} \frac{1}{1-p^{-s}} = \sum_{\substack{n=1 \\ n \text{ has no prime factors} \\ \text{exceeding } x}}^{\infty} \frac{1}{n^s}$$

Letting x tend to infinity completes the proof of Euler's formula. □

The first thing Euler's did with his product formula is to again show the infinity of primes.

Proof: To be slightly more rigorous than Euler, we instead take the truncated formula with $s = 1$:

$$\sum_{n=1}^{\infty}{}' \frac{1}{n} = \prod_{p \in \mathbb{P}_x} \frac{1}{1 - 1/p}.$$

where the $'$ denotes the restriction "has no prime factors exceeding x". Letting $x \to \infty$ on the left yields the divergent harmonic series. But if $\mathbb{P}$ were finite, say $\mathbb{P} = \{p_1, \ldots, p_k\}$, then the same limit on the right would be finite:

$$\prod_{p \in \mathbb{P}} \frac{1}{1 - 1/p} = \frac{1}{1 - 1/p_1} \cdots \frac{1}{1 - 1/p_k} < \infty.$$

Hence $\mathbb{P}$ cannot be finite. $\square$

We can now answer Euler's original question.

$$\sum_{p\in\mathbb{P}} \frac{1}{p} \quad \text{diverges}$$

Proof: Again we start with

$$\sum_{n=1}^{\infty}{}' \frac{1}{n} = \prod_{p\in\mathbb{P}_x} \frac{1}{1-1/p}.$$

Taking the logarithm this yields

$$\log\left(\sum_{n=1}^{\infty}{}' \frac{1}{n}\right) = \log\left(\prod_{p\in\mathbb{P}_x} \frac{1}{1-1/p}\right)$$
$$= \sum_{p\in\mathbb{P}_x} \log\left(\frac{p}{p-1}\right) = \sum_{p\in\mathbb{P}_x} \log\left(1+\frac{1}{p-1}\right).$$

The next step is to recall the inequality

$$\log(1+x) < x.$$

Therefore

$$\sum_{p\in\mathbb{P}_x} \log\left(1+\frac{1}{p-1}\right) < \sum_{p\in\mathbb{P}_x} \frac{1}{p-1},$$

which in its turn implies that

$$\log\left(\sum_{n=1}^{\infty}{}' \frac{1}{n}\right) < \sum_{p\in\mathbb{P}_x} \frac{1}{p-1} = \sum_{i=1}^{m} \frac{1}{p_i - 1},$$

where p_i is the ith prime and, as before, $m = |\mathbb{P}_x|$.

Finally note that for $i \geq 2$ there holds $\dfrac{1}{p_i - 1} \leq \dfrac{1}{p_{i-1}}$, so that

$$\log\left(\sum_{n=1}^{\infty}{}' \frac{1}{n}\right) < 1 + \sum_{i=2}^{m} \frac{1}{p_{i-1}} = 1 + \sum_{i=1}^{m-1} \frac{1}{p_i} < 1 + \sum_{p\in\mathbb{P}_x} \frac{1}{p}.$$

The inequality

$$\sum_{p\in\mathbb{P}_x} \frac{1}{p} > -1 + \log\left(\sum_{n=1}^{\infty}{}' \frac{1}{n}\right)$$

implies that

$$\sum_{p\in\mathbb{P}} \frac{1}{p}$$

diverges. Indeed, taking the limit $x \to \infty$ the right-hand side effectively becomes the logarithm of the divergent harmonic series. □

Using more sophisticated methods it may be shown that

$$\lim_{x\to\infty}\left(\sum_{p\in P_x} \frac{1}{p} - \log(\log x)\right) = 0.2614972128476427837554268386\ldots$$

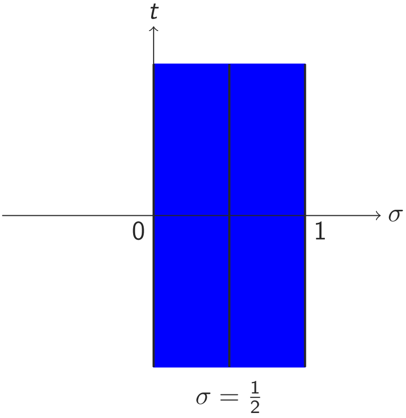
where the constant on the right is known as the Meissel–Mertens constant.

The Riemann zeros

In 1896 the PNT was independently proved by Jacques Salomon Hadamard and Charles-Jean Étienne Gustave Nicolas Levieux (Baron) de la Vallée–Poussin.

Their proofs are much too advanced to reproduce here, and are related to the zeros of the Riemann zeta function.

Specifically, they showed that $\zeta(s)$ has no zeros on the boundary of what is known as the critical strip. This fact is equivalent to the PNT, and clearly demonstrates the importance of understanding the (location of the) zeros of the Riemann zeta function.



The critical strip of the Riemann zeta function in the complex s-plane, with $s = \sigma + \mathrm{i}t$.

The line $\sigma = \frac{1}{2}$ is know as the critical line.

Since 1896 a number of "elementary" proofs of the PNT have been found. The most famous and controversial being the proof(s) of Atle Selberg and Paul Erdős, which does not (do not) require complex analysis.

Their proof(s) rest on the following key inequality due to Selberg:

$$\sum_{\substack{p\in\mathbb{P}\\ p<x}} \log^2 p + \sum_{\substack{p,q\in\mathbb{P}\\ pq<x}} \log p \, \log q = 2x \log x + O(x).$$

Here the big O notation $f(x) = O(g(x))$ means there exists an x_0 such that

$$|f(x)| \leq M|g(x)| \quad \text{for all } x > x_0.$$

Let Ω be an open subset of $\mathbb{C}$. The function $f : \Omega \mapsto \mathbb{C}$ is analytic in $z_0 \in \Omega$ if there exists a neighbourhood of z_0 such that f can be represented as a power series around z_0:

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

If f is analytic at every point in Ω then f is said to be analytic on Ω.

Examples of analytic functions on $\mathbb{C}$ are polynomials, $\sin z$, $\cos z$ and e^z.

Remark: Often the terms analytic and holomorphic are used interchangeably. The latter roughly refers to being complex differentiable, i.e., the existence of

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

for all $z_0 \in \Omega$. It is one of the main theorems of complex analysis that analyticity and holomorphicity are one and the same.

It often occurs in mathematics that one is given a function that is analytic on some $\Omega \subset \mathbb{C}$ and we want to extend that function to an analytic function on a larger domain $\Omega' \subset \mathbb{C}$. The successful construction of such an extension goes by the name of analytic continuation.

More precisely, if $\Omega \subset \Omega'$ are open subsets of $\mathbb{C}$ and $f : \Omega \mapsto \mathbb{C}$ and $g : \Omega' \mapsto \mathbb{C}$ are analytic functions on Ω and Ω', respectively, such that

$$g|_{\Omega} = f$$

(read: g restricted to Ω is f) then g is called the analytic continuation of f from Ω to Ω'.

Note: If the analytic continuation of f from Ω to Ω' exists then it is unique.

As a simple example take $\Omega = \{z \in \mathbb{C} : \ |z| < 1\}$ and $\Omega' = \mathbb{C} \setminus \{1\}$ and define $f : \Omega \mapsto \mathbb{C}$ and $g : \Omega' \mapsto \mathbb{C}$ by

$$f(z) = \sum_{n=0}^{\infty} z^n$$

and

$$g(z) = \frac{1}{1-z}.$$

Then both f and g are analytic functions on their respective domains and

$$g|_{\Omega} = f.$$

Hence g is the analytic continuation of f to Ω'.

The previous example is typical. The reason a particular representation of a function, such as $\sum_n z^n$, defines an analytic function on Ω and not on $\Omega' \supset \Omega$ often is the occurrence of a pole on $\partial\Omega$, the boundary of Ω.

As an alternative to saying that

$$\frac{1}{1-z}$$

is the analytic continuation of

$$\sum_{n=0}^{\infty} z^n$$

from $|z| < 1$ to $\mathbb{C} \setminus \{1\}$ one can say that the former is the analytic continuation of the latter from $|z| < 1$ to a meromorphic function on $\mathbb{C}$ with a (simple) pole at $z = 1$.

A not-so-simple example of analytic continuation concerns the Euler Gamma function, a function needed in order to formulate the Riemann hypothesis.

The Gamma function arises as a generalisation of the well-known $n!$:

$$\Gamma(z) = \int_0^\infty x^{z-1} \mathrm{e}^{-x} \mathrm{d}x, \qquad \mathrm{Re}(z) > 0.$$

Note that integration by parts (with $f = x^n$ and $g' = \mathrm{e}^{-x}$) gives

$$\begin{aligned} \Gamma(n+1) &= \int_0^\infty x^n \mathrm{e}^{-x} \mathrm{d}x \\ &= -\Big[x^n \mathrm{e}^{-x}\Big]_0^\infty + n \int_0^\infty x^{n-1} \mathrm{e}^{-x} \mathrm{d}x \\ &= n\,\Gamma(n) \end{aligned}$$

for $n \geq 1$. Since $\Gamma(1) = \displaystyle\int_0^\infty \mathrm{e}^{-x} \mathrm{d}x = 1$ it follows that $\Gamma(n+1) = n!$.

The integral defining the Gamma function is absolute convergent for $\mathrm{Re}(z) > 0$ and defines an analytic function for $\mathrm{Re}(z) > 0$.

Question: Can we continue $\Gamma(z)$ to all of $\mathbb{C}$?

The answer is affirmative and we give two approaches (both of which will play a role later when we discuss the Riemann zeta function).

Method 1: The reason the integral representation of the Gamma function causes problems for $\mathrm{Re}(z) \leq 0$ is because for $z = 0$ and small x the integrand behaves like $1/x$ which cannot be integrated from 0 to ϵ. Hence we split the integral into a "good" and a "bad" part:

$$\begin{aligned}\Gamma(z) &= \int_0^\infty x^{z-1} \mathrm{e}^{-x} \mathrm{d}x \\ &= \int_0^1 x^{z-1} \mathrm{e}^{-x} \mathrm{d}x + \int_1^\infty x^{z-1} \mathrm{e}^{-x} \mathrm{d}x.\end{aligned}$$

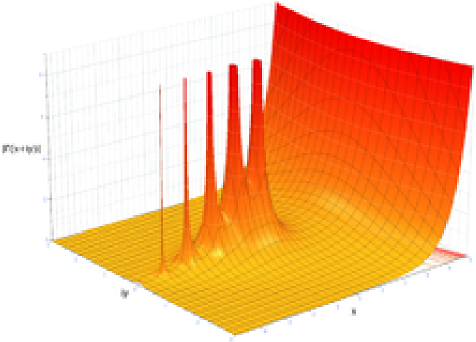
The good part is an entire function (analytic/holomorphic on all of $\mathbb{C}$). For the bad part we use the series representation of the exponential function:

$$e^{-x} = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!}.$$

Interchanging integral and sum (this can be justified) gives

$$\begin{aligned}\Gamma(z) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^1 x^{n+z-1} dx + \int_1^{\infty} x^{z-1} e^{-x} dx \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+z)n!} + \int_1^{\infty} x^{z-1} e^{-x} dx.\end{aligned}$$

The right-hand side is a meromorphic function on $\mathbb{C}$ with simple poles at the non-positive integers.



The absolute value of the Gamma function $\Gamma(z)$ in the $z = x + iy$ plane. The spikes represent the simple poles at the non-positive integers.

Method 2: Carrying out integration by parts exactly as before implies the functional equation $\Gamma(z+1) = z\,\Gamma(z)$, which we write as

$$\Gamma(z) = \frac{\Gamma(z+1)}{z}.$$

But the right hand side defines an analytic function on $\mathrm{Re}(z) > -1$ except for $z = 0$ where it has a simple pole, i.e., it defines a meromorphic function on $\mathrm{Re}(z) > -1$ with a simple poles at $z = 0$.

The functional equation can obviously be iterated. For example,

$$\Gamma(z) = \frac{\Gamma(z+2)}{z(z+1)}.$$

with on the right-hand side a meromorphic function on $\mathrm{Re}(z) > -2$ with simple poles at $z = 0$ and $z = -1$, etc.

Summary: One can use the functional equation

$$\Gamma(z+1) = z\,\Gamma(z)$$

to analytically continue $\Gamma(z)$ from $\mathrm{Re}(z) > 0$ to a meromorphic function on $\mathbb{C}$ with simple poles at the non-positive integers.

With a little more work it may also be shown that the Gamma function satisfies the Euler reflection formula

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}$$

We will establish something quite similar for the Riemann zeta function.

As a final result for the Gamma function we state the product formula for its reciprocal:

$$\frac{1}{\Gamma(z)} = z e^{\gamma z} \prod_{n=1}^{\infty} \left\{ \left(1 + \frac{z}{n}\right) e^{-nx} \right\},$$

where γ is the The Euler–Mascheroni constant

$$\gamma = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \log n \right) = 0.5772156649015328606065120900 82\ldots$$

The product formula shows that $1/\Gamma(z)$ is an entire function so that

$\Gamma(z)$ has no zeros

The definition of the Riemann zeta function as an infinite series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

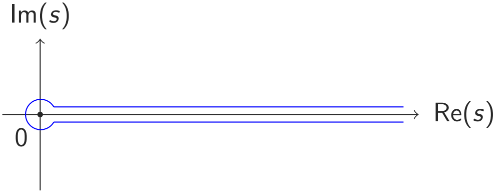
implies that $\zeta(s)$ is analytic for $\mathrm{Re}(s) > 1$.

In his study of the zeta function, resulting in his famous hypothesis, Bernhard Riemann used contour integration to continue $\zeta(s)$ to a meromorphic function on $\mathbb{C}$ with a single simple pole located at $s = 1$.

Riemann's contour integral representing the zeta function is

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi \mathrm{i}} \int_{C} \frac{(-x)^s}{\mathrm{e}^x - 1} \cdot \frac{\mathrm{d}x}{x},$$

where C is the positively oriented contour



To avoid contour integration we will follow a different route, reminiscent of the first method for continuing the Gamma function.

The reason the series representation of the zeta function defines an analytic function for $\mathrm{Re}(s) > 1$ only is because of the divergence of the harmonic series, i.e., "$\zeta(1) = \infty$".

Based on the previous examples of the geometric series and the Gamma function we suspect a pole at $s = 1$, and as before, we would like to split the zeta function into a into a "good" and a "bad" part.
This can be done as follows:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = s \sum_{n=1}^{\infty} \int_n^{\infty} \frac{\mathrm{d}x}{x^{s+1}}.$$

Now let $\lfloor x \rfloor$ and $\{x\}$ denote the integer part and fractional part of x.

Then

$$\begin{aligned}\sum_{n=1}^{\infty} \int_{n}^{\infty} f(x)\mathrm{d}x &= \sum_{n=1}^{\infty} n \int_{n}^{n+1} f(x)\mathrm{d}x \\ &= \sum_{n=1}^{\infty} \int_{n}^{n+1} \lfloor x \rfloor f(x)\mathrm{d}x = \int_{1}^{\infty} \lfloor x \rfloor f(x)\mathrm{d}x.\end{aligned}$$

Hence

$$\begin{aligned}\zeta(s) &= s \int_{1}^{\infty} \frac{\lfloor x \rfloor}{x^{s+1}} \, \mathrm{d}x \\ &= s \int_{1}^{\infty} \frac{x - \{x\}}{x^{s+1}} \, \mathrm{d}x = \frac{s}{s-1} - s \int_{1}^{\infty} \frac{\{x\}}{x^{s+1}} \, \mathrm{d}x.\end{aligned}$$

The integral

$$\int_1^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x$$

converges (absolutely) for $\mathrm{Re}(x) > 0$ and defines an analytic function on this domain.
Indeed, for $s = \sigma + \mathrm{i}t$,

$$\left|\int_1^L \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x\right| \leq \int_1^L \frac{\{x\}}{x^{\sigma+1}}\,\mathrm{d}x \leq \int_1^L \frac{\mathrm{d}x}{x^{\sigma+1}} = \frac{1}{\sigma}\left[1 - \frac{1}{L^\sigma}\right].$$

In the limit $L \to \infty$ the right converges to $1/\sigma$ provided that $\sigma > 0$.

The expression

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x$$

thus provides the analytic continuation of $\zeta(s)$ from $\mathrm{Re}(s) > 1$ to a meromorphic function on $\mathrm{Re}(s) > 0$ with a simple pole at $s = 1$.

As before, let $s = \sigma + \mathrm{i}t$.
The previous manipulations have extended the domain of the Riemann zeta function from $\sigma > 1$ to $\sigma > 0$.

To extend this even further to all of σ (i.e., all of the complex s-plane) we need something akin to the "second method" employed in the continuation of the Gamma function.

The key is to prove (we will only sketch a proof) that for $0 < \sigma \leq 1$ the zeta function satisfies the functional equation

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{1}{2}\pi s\right)\Gamma(1-s)\zeta(1-s).$$

which we will write as

$$\zeta(s) = \chi(s)\zeta(1-s)$$

with

$$\chi(s) = 2^s \pi^{s-1} \sin\left(\frac{1}{2}\pi s\right)\Gamma(1-s).$$

Let us first consider the function $\chi(s)$.

Since $\Gamma(s)$ has a simple pole at every non-positive integer, $\Gamma(1-s)$ has a simple pole at every positive integer. But

$$\sin\left(\frac{1}{2}\pi s\right) = 0$$

for $s \in 2\mathbb{Z}$ so that $\chi(s)$ is a meromorphic function on $\mathbb{C}$ whose only poles are the positive odd integers, and whose only zeros are the non-positive even integers:

$$\chi(s) \text{ has } \begin{cases} \text{poles at } s = 1, 3, 5, \ldots \\ \text{zeros at } s = 0, -2, -4, \ldots \end{cases}$$

Note: The above-listed poles are simple and the zeros have multiplicity 1.

If we replace $s \mapsto 1 - s$ in the functional equation we get

$$\zeta(s) = \chi(s)\zeta(1-s) \quad \mapsto \quad \zeta(1-s) = \chi(1-s)\zeta(s)$$

so that we must have

$$\chi(s)\chi(1-s) = 1.$$

Using the explicit formula for $\chi(s)$ and the trigonometric identities $\sin(\pi/2 - x) = \cos x$ and $2 \sin x \cos x = \sin(2x)$ this is easily seen to be equivalent to the Euler reflection formula

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s}.$$

Now what is the point of the functional equation

$$\zeta(s) = \chi(s)\zeta(1-s)?$$

To answer this question, note that up to an easy factor $\chi(s)$ it maps the zeta function for $\sigma > 1/2$ onto the zeta function for $\sigma < 1/2$. Since we already have a representation of the zeta function for all of $\sigma > 0$ we can use the functional equation to analytically continue to all of the complex s-plane. In doing so we can forget about the fact that initially the functional equation was proved (of course we haven't yet) for $0 < \sigma \leq 1$.

In other words, we define the analytic continuation $\boldsymbol{\zeta}(s)$ of the function $\zeta(s)$ by

$$\boldsymbol{\zeta}(s) = \begin{cases} \dfrac{s}{s-1} - s\displaystyle\int_1^\infty \dfrac{\{x\}}{x^{s+1}}\,\mathrm{d}x & \text{for } \sigma > 0, \\ \chi(s)\zeta(1-s) & \text{for } \sigma \leq 0. \end{cases}$$

For the time being, let us accept the functional equation and the above continuation of the zeta function to all of the complex s-plane.

What can we then say about the zeros and poles of $\zeta(s)$, and how does this relate to the distribution of the primes?

First we will answer the question concerning the poles of $\zeta(s)$.

We already know that for $\sigma > 0$ the zeta function $\zeta(s)$ has a simple pole at $s = 1$. But for $\sigma \leq 0$ we have $\zeta(s) = \chi(s)\zeta(1-s)$.
Now $\chi(s)$ only has poles at the positive odd integers so is free of poles for $\sigma \leq 0$.
The function $\zeta(1-s)$ has a simple pole but $\chi(s)$ has a zero at $s = 0$ (at every non-positive even integer in fact) so that their product does not have a pole at $s = 0$. Hence

$\zeta(s)$ has a single (simple) pole at $s = 1$

Next we consider the question concerning the zeros of $\boldsymbol{\zeta}(s)$.

Recall Euler's product formula

$$\zeta(s) = \prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}} \qquad \text{for } \sigma > 1.$$

Since $1/(1 - p^{-s}) \neq 0$ it is clear that

$$\zeta(s) \neq 0 \qquad \text{for } \sigma > 1.$$

We can use the above to find the zeros of $\boldsymbol{\zeta}(s)$ for $\sigma < 0$. Indeed, since $\boldsymbol{\zeta}(s) = \chi(s)\zeta(1 - s)$ and since $\zeta(s)$ has no zeros for $\sigma > 1$ (so that $\zeta(1 - s)$ has no zeros for $\sigma < 0$) the only potential zeros of $\boldsymbol{\zeta}(s)$ for $\sigma < 0$ correspond to the zeros of $\chi(s)$.
The zeros of the latter are at $s = 0, -2, -4, \dots$. The point $s = 0$ is not a zero of $\boldsymbol{\zeta}(s) = \chi(s)\zeta(1 - s)$ however, since the zero (of multiplicity 1) of $\chi(s)$ is multiplied by the simple pole of $\zeta(1 - s)$ at $s = 0$.

$\zeta(s)$ has zeros at $s = -2, -4, \ldots$.

The above zeros of the zeta function are known as the trivial zeros.

Hence what we have established so far is that away from the critical strip, i.e., for $\sigma > 1$ or $\sigma < 0$ the only zeros of $\zeta(s)$ are the trivial zeros.

The following theorem, which implies the PNT was proved by Hadamard and de la Vallée–Poussin:

$\zeta(s)$ has no zeros on the boundary of the critical strip, i.e., $\zeta(s) \neq 0$ for $\sigma = 1$

The famous Riemann hypothesis sharpens the previous statement considerably

All zeros of $\zeta(s)$ lie on the critical line $\sigma = \frac{1}{2}$

Unsurprisingly, this implies a sharpening of the PNT, and in 1901 Niels Fabian Helge von Koch proved that

If the Riemann hypothesis is true then

$$\pi(x) = \mathrm{Li}(x) + O(x^{1/2} \log x)$$

x	$\pi(x)$	$\left[\pi(x) - \mathrm{Li}(x)\right] / (x^{1/2} \log x)$
10^3	168	-0.039207105082948185470
10^6	78,498	-0.009301429371206646721
10^9	50,847,534	-0.002593989418422588235
10^{12}	37,607,912,018	-0.001384739258863384220
10^{15}	29,844,570,422,669	-0.000963748362252954309
10^{18}	24,739,954,287,740,860	-0.000529587231803321958

We now come to the "proof" of the functional equation

$$\zeta(s) = \chi(s)\zeta(1-s),$$

where

$$\chi(s) = 2^s \pi^{s-1} \sin\left(\frac{1}{2}\pi s\right)\Gamma(1-s).$$

Proof: As a first step we take the definition of the Gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx, \qquad \mathrm{Re}(z) > 0,$$

replace $z \mapsto s/2$, and make the change of integration variable $x \mapsto \pi n^2 t$:

$$\Gamma\left(\frac{s}{2}\right)\frac{1}{\pi^{s/2} n^s} = \int_0^\infty t^{s/2-1} e^{-\pi n^2 t} dt.$$

Next we sum n over the positive integers, resulting in

$$\Gamma\left(\frac{s}{2}\right)\frac{\zeta(s)}{\pi^{s/2}} = \int_0^\infty t^{s/2-1}\sum_{n=1}^\infty e^{-\pi n^2 t}dt = \int_0^\infty t^{s/2-1}\,\Theta(t)dt,$$

where $\Theta(t) := \sum_{n=1}^\infty e^{-\pi n^2 t}$.

Following earlier manipulations for the Gamma function we split the integral into two parts:

$$\Gamma\left(\frac{s}{2}\right)\frac{\zeta(s)}{\pi^{s/2}} = \left[\int_0^1 + \int_1^\infty\right] t^{s/2-1}\,\Theta(t)\,dt.$$

The idea is now to rewrite the integral over $[0,1]$ as an integral over $[1,\infty)$ using the variable change $t \mapsto 1/t$.

Since

$$t \mapsto 1/t \quad \text{implies} \quad \mathrm{d}t \mapsto -\mathrm{d}t/t^2,$$

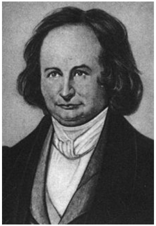
we get

$$\int_0^1 t^{s/2-1}\,\Theta(t)\,\mathrm{d}t = -\int_\infty^1 t^{-s/2-1}\,\Theta(1/t)\,\mathrm{d}t$$
$$= \int_1^\infty t^{-s/2-1}\,\Theta(1/t)\,\mathrm{d}t.$$

The Jacobi theta function $\vartheta(z)$ is defined as

$$\vartheta(z) = \sum_{n=-\infty}^{\infty} \mathrm{e}^{-\pi n^2 z}.$$

This function is the fundamental building block in the theory of elliptic functions.



The Jacobi theta function satisfies the fundamental transformation property ("proved" later):

$$\vartheta(1/z) = \sqrt{z}\,\vartheta(z)$$

Since

$$\vartheta(z) = \sum_{n=-\infty}^{\infty} \mathrm{e}^{-\pi n^2 z} = 1 + 2\sum_{n=1}^{\infty} \mathrm{e}^{-\pi n^2 z} = 1 + 2\Theta(t).$$

this implies a similar, if somewhat less elegant transformation for $\Theta(z)$:

$$\Theta(1/z) = \frac{\sqrt{z}-1}{2} + \sqrt{z}\,\Theta(z).$$

Hence

$$\int_0^1 t^{s/2-1}\,\Theta(1/t)\,\mathrm{d}t = \frac{1}{s(s-1)} + \int_1^{\infty} t^{-s/2-1/2}\,\Theta(t)\,\mathrm{d}t.$$

As a consequence of this integral transformation for $\Theta(t)$ we arrive at

$$\Gamma\left(\frac{s}{2}\right)\frac{\zeta(s)}{\pi^{s/2}} = \frac{1}{s(s-1)} + \int_1^\infty \left[t^{s/2-1} + t^{-s/2-1/2}\right]\Theta(t)\,\mathrm{d}t.$$

Recall that $\Theta(t) = \sum_{n=1}^\infty \mathrm{e}^{-\pi n^2 t} = O(\mathrm{e}^{-\pi t})$. This exponential decay of the integrand implies uniform convergence of the integral for all $s \in \mathbb{C}$, which in its turn implies that the integral on the right defines an analytic function on $\mathbb{C}$.

But next observe that the right-hand side is invariant under the change

$$s \mapsto 1 - s\,!$$

The right-hand side must obviously satisfy this same invariance, so that

$$\Gamma\left(\frac{s}{2}\right)\frac{\zeta(s)}{\pi^{s/2}} = \Gamma\left(\frac{1-s}{2}\right)\frac{\zeta(1-s)}{\pi^{(1-s)/2}},$$

or

$$\zeta(s) = \chi(s)\zeta(1-s) \quad \text{where} \quad \chi(s) = \pi^{s-1/2}\,\frac{\Gamma\left(\frac{1-s}{2}\right)}{\Gamma\left(\frac{s}{2}\right)}.$$

But

$$\chi(s) = \pi^{s-1/2}\,\frac{\Gamma\left(\frac{1-s}{2}\right)}{\Gamma\left(\frac{s}{2}\right)}$$
$$= \pi^{s-1/2}\,\frac{\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(\frac{1-s}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{s}{2}\right)\Gamma\left(1-\frac{s}{2}\right)}$$
$$= \pi^{s-3/2}\,\sin\left(\frac{1}{2}\pi s\right)\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(\frac{1-s}{2}+\frac{1}{2}\right),$$

where the last equality follows thanks to the reflection formula for the Gamma function.
Finally employing Legendre's duplication formula

$$\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) = \sqrt{\pi}\,2^{1-2z}\Gamma(2z)$$

shows that

$$\chi(s) = 2^s\pi^{s-1}\sin\left(\frac{1}{2}\pi s\right)\Gamma(1-s).$$

□

It finally remains to be proven that the Jacobi theta function transforms as

$$\vartheta(1/z) = \sqrt{z}\,\vartheta(z)$$

This is an immediate consequence of classical result in Fourier theory known as the Poisson summation formula:

Subject to some (mild) conditions

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

where

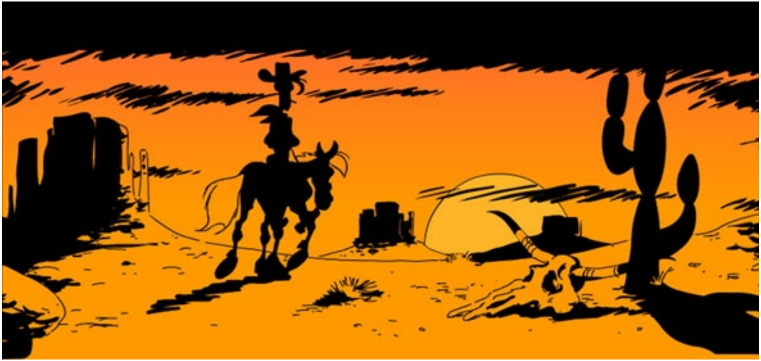
$$\hat{f}(n) = \int_{-\infty}^{\infty} f(x) e^{-2\pi n i x} dx$$

Taking $f(x) = \mathrm{e}^{-\pi a x^2}$ gives

$$\begin{aligned}\hat{f}(n) &= \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi n \mathrm{i} x - \pi a x^2} \mathrm{d}x \\ &= \mathrm{e}^{\pi n^2/a} \int_{-\infty}^{\infty} \mathrm{e}^{-\pi a (x + n\mathrm{i}/a)^2} \mathrm{d}x \\ &= \frac{\mathrm{e}^{\pi n^2/a}}{\sqrt{a}},\end{aligned}$$

where the last equality follows from the standard Gaussian integral.

By the Poisson summation formula this results in the transformation formula for the Jacobi theta function.



The End