Appendix 9

A proof of Riemann’s hypothesis via Denjoy’s equivalent theorem

On 2006 06 27 I announced on the net a theorem very much stronger than Riemann’s. Write $\mathrm{R}_2(n)$ for $\mathrm{li}\,n$ -1/2$\mathrm{li}\,n^{1/2}$. Then for all $n>1$

(1) $\mathrm{R}_2(n) - 1/2(\mathrm{R}_2(n))^{1/2} < \pi(n) < \mathrm{R}_2(n) + 1/2(\mathrm{R}_2(n))^{1/2}$

My theorem in (1) is about as much stronger than the RH as the RH is stronger than the PNT. In working up a more-general account of this theorem, with rigorous proofs of its validity, for eventual publication in print, I discovered a neat proof, this time of Riemann’s hypothesis only, on entirely different lines.

For this second proof we refer to what is called Denjoy’s probabilistic interpretation, notably that the RH is equivalent to the proposition that any square-free number, taken at random, has an equal probability of containing an odd or an even number of distinct prime divisors.

Legendre's formula for $\pi(n)$

Legendre's formula (*Essai* 2nd edition, Paris 1808, pp 412sq) is a recipe for calculating the exact number $\pi(n)$ of primes $\leq n$ without identifying them all. It can be written

(2) $\pi(n) = \pi(n^{1/2}) + (\Sigma\, \mu(d)\ [n/d]) - 1$

where μ is the Möbius function, and the denominators (d) are all the natural numbers that have no large prime p in their decomposition. A prime p is ‘large’ (in relation to n) if $p > n^{1/2}$.

Analysis. The formula in (2) works correctly because its summation term yields the number of numbers $\leq n$ that are not struck out by the Eratosthenes procedure of striking out those of them that are divisible by a prime q that is ‘small’ in relation to n, i.e. is such that $2 \leq q \leq n^{1/2}$. The unstruck numbers include 1, which is not nowadays classed as a prime. Students of arithmetic born before 1900 were taught that 1 is the least prime, making Goldbach’s conjecture apply to all even numbers including 2. Present-day arithmeticians find it more convenient to exclude 1 from the class of prime numbers, making it the unique natural number whose number of distinct prime divisors is zero.

The function $\mu(d)$ can now be defined as equal to 1 if the number of distinct prime divisors of d is even, –1 if it is odd, and 0 if d has a repeated divisor (other than 1). Since 1 is not struck out by the sieve of Erastosthenes and is also included in the count of "primes" calculated by the section $\Sigma\mu(d)\ [n/d]$[*], the count must be reduced by one in either case, and then to get the complete answer the number of small primes (q) used as strikers must be added to the total.

Illustration of Legendre's formula with $n = 20$

d	$f(d) = \mu(d)\ [n/d]$
1*	+20
2*	−10
3*	−6
5	−4
6*	+3
7	−2
10	+2
11	−1
13	−1
14	+1
15	+1
17	−1
19	−1
	Σ +1

$\Sigma\mu(d^*)\ [n/d^*] = 7$

$7 - 1 + \pi(n^{1/2}) = 8 = \pi(20)$

The (d*) are the denominators with no large prime in their decomposition. The small primes 2, 3 must be known explicitly, then the number of large primes 5, 7, 11, 13, 17, 19 can be calculated without any of them being identified.

In the table above we see an illustration of the use of Legendre's formula to calculate $\pi(n)$ for $n = 20$. The starred terms, with no large prime divisors of d, are used to calculate the number of large primes $\leq n$. Notice I have used all the (d) that yield an $f(d)$ other than zero, and the sum of these, for any n, must always be 1, since only one number, 1 itself, remains unstruck if we use all the primes.

The stage is now set for my proof of Denjoy's equivalent to Riemann's hypothesis.

First we get rid of the 1, which is the only number left standing after my extension of Legendre's procedure. To do this we remove it at the beginning, quite legitimately, because it is neither prime nor composite, and so does not belong to either of the two complementary classes, composites and primes, to which we reduce the number system in accordance with modern practice.

We thus further rectify the procedure by making the following change:

use $f(d) = \mu(d)\ [(n-1)/d]$ for $d=1$

and use $f(d) = \mu(d)\ [n/d]$ for all other values of d.

---

[*] The fact that, with d unrestricted, the formula $\Sigma\mu(d)\ [n/d] = 1$ is true for all n was first noted by Meissel in *Observationem quaedam in theoria numerorum*, Berlin 1850, but he failed to discover my easy proof of it (below) or to find a use for it.

Now rework n=20 using the new procedure

d	$f(d)$		
1	+19		
2	−10		
3	−6		upper section
5	−4		
6	+3		
7	−2	sum to half way	
10	+2	+2	
11	−1		
13	−1		
14	+1		lower section
15	+1		
17	−1	sum in 2nd half of n	
19	−1	−2	
	0	sum complete.	

I have divided the terms into two sections, in the upper of which each $f(d)$ consists of $\mu(d)$ multiplied by some positive number >1, and in the lower each $f(d)$ is simply $\mu(d)$. (We should note that every value of $\mu(d)$ other than the first must appear in the lower section of one or more natural n.) In the upper section the $f(d)$ sum need not be exactly the sum of the $\mu(d)$ pluses and minuses to this point, but it is obviously positively correlated to it. (For example if all the terms were negative the answer would be negative, and vice versa.) In the lower section the sum of the terms is exactly the sum of the $\mu(d)$ in the section.

Recall that, by Denjoy's equivalent, the Riemann hypothesis is true if and only if the algebraic sums of the pluses and minuses of $\mu(d)$, taken progressively at unit increments of d, vary asymptotically around zero[*]. Suppose it is untrue. This can only mean that the sums must vary asymptotically around some number other than zero[**].

Suppose this number is such that the upper sections of all numbers, taken progressively, vary in aggregate around +2, which we recognize is the sum of the upper-section terms for $n = 20$, so we may take this n as a typical example. Now the sum of the lower-section terms for this n must be exactly −2 to compensate the upper section. So double n *to* $2n = 40$. The aggregate of pluses and minuses in the upper section of $2n = 40$ is exactly what it was in the whole of $n = 20$. But it contains two more minus signs than

---

[*] This means we can get it as close as we like to an average of zero difference between the two.

[**] This means we cannot get the average difference as close as we like to zero after n has reached a certain size, but can get it as close as we like to some other number.

did the upper section of $n = 20$, so its sum is likely to be reduced towards or beyond zero. Suppose by an unlikely chance it is still +2. The lower section of $2n = 40$ must again be –2 to compensate this, so repeat the procedure by doubling $2n$ to $4n = 80$. Now the upper section of this new number $4n = 80$ must contain two more extra minus signs, making it even more likely to be reduced towards or beyond zero.

These unlikely chances cannot continue for ever, because every time we doubled the argument we would have to add an average of two more minus signs to the aggregate of the upper-section terms of the new doubled argument, so there must come a time when the aggregate of the upper-section terms of the new doubled argument is reduced to or beyond zero. Suppose it is reduced to zero. Then the aggregate of the lower-section terms for this argument will also be zero, and there will be no tendency in either direction when it is doubled again. But suppose the aggregate in the upper section is reduced beyond zero to a negative value. Now the lower-section aggregate for this argument must be positive, and the whole process must play itself out again, this time in the opposite direction.

Recall, finally, that the lower sections of every argument consist of increasingly protracted sets of consecutive terms of $\mu(d)$, and that all values of $\mu(d)$ except the first must be presented in the lower sections of one or more natural n.

Because of the negative feedback between the two sections, to suppose the aggregate of their pluses and minuses can vary around any number other than zero is self-contradictory. Therefore it varies around zero and the Riemann hypothesis is true.

Telephone +44 1985 844 855

Aftermath

The process of negative feedback in successive Möbius values is continuous, and begins long before the big jumps I used to illustrate it. Those of my readers who remember their course in radio telephony will recall that negative feedback leads to oscillation (called 'hunting') about a mean value. The negative feedback in the case I considered is quite pronounced, with an average gain of 1.7857 in the early stages, as the following table will confirm.

n	upper-section deviation	added	swing in direction of addition
5	+2	–	–
10	–1	–2	–3
20	+2	+1	+3
40	–3	–2	–5
80	+4	+3	+7
160	–2	–4	–6
320	–1	+2	+1
		absolute sum 14	absolute sum 25

average gain = 1.785714...

From the table we can see that the successive Möbius values are not at all random, as many commentators have mistakenly supposed, but follow what is called by unsophisticated gamblers a "maturity of chances" hypothesis. This is a supposition that, for example, after a run of successive black numbers at the roulette table, the probability of a red number appearing next is increased to "redress the balance". But the (to some extent) empirically verified hypothesis of probability determines axiomatically that the result of the next spin will be independent of previous results, so the player loses from the inclusion of a zero number that renders the probability of red or black slightly less than 1/2. But if the casino were naïve enough to offer evens against a + or – appearing in any continous set of consecutive values of $\mu(n)$, the player could win a fortune by betting against the trend.

Offhand I can think of no series of naturally-produced numbers, other than $\mu(n)$, for which a maturity-of-chances hypothesis happens to be true.* Can any of my readers?

An unintended confirmation of this is provided by John Derbyshire in his book *Prime obsession* (New York 2003), which is the best and most complete account of the Riemann hypothesis I have seen. It contains, moreover, fewer serious mistakes than any other

---

* Of course it is also true of primes and composites, but these categories are not independent of $\mu(n)$.

account I have read, though it is of course impossible to write a book of this size (422 pages) without including some mistakes. He makes the common mistake of suggesting that a series of consecutive values of $\mu(n)$ might be random in respect of their + and – signs (p 322). In a previous page (250) he quotes sets of Mertens's function (cumulative $\mu(n)$) that he says tell us very little except that their absolute value increases as n does. In fact they tell us a great deal, and if he had noticed it he might, admittedly with a fair amount of further detective work, have discovered this beautiful proof of Riemann's hypothesis several years before I did.

On page 322 he correctly points out that the average difference between n randomly-produced +1's and –1's is $\sqrt{n}$. To convert Mertens's function into a corresponding function for square-free (n) we must multiply each n by about 0.608. From his second set of figures we find arguments

1000 2000 3000 4000 5000 6000 7000 8000 9000 10000

conversions

608 1216 1824 2432 3040 3648 4256 4864 5472 6080

roots of conversions

25 35 43 49 55 60 65 70 74 78

Mertens's values for original arguments

2 5 –6 –9 2 0 –25 –1 1 –23.

It is evident that the final set of values (considered absolutely as differences) are ridiculously below what they should be if the original sets of +1's and –1's were randomly produced. I was going to do a table of his third set of arguments, in millions, whose values are equally impressive, but the set I have tabulated all denote unrandomness so obviously that I will leave the tabulation of his third set to the reader, for the good feeling of being part of the research. What they show is that the successive Möbius +1's and –1's are unrandom to an enormous degree, being hugely biased towards a maturity-of-chances hypothesis.

My strong theorem at the beginning of this memoir shows that the primes are similarly unrandom, in the sense of being much more evenly-spread than we thought they might be.

I had experimented with Legendre's method of counting primes for many years, convinced that it could lead to a very elementary proof of the PNT, but ironically could not see how to do it until faced with the more heroic prospect of proving the Riemann hypothesis. In this case it seems to be impossible to prove the one without simultaneously proving the other.

In common with other experts in the field, I had begun to suspect that the RH is a problem in elementary arithmetic and not in analysis, as many of us, probably including Riemann, had previously thought.

Telephone +44 1985 844 855

Abbreviations

PNT = prime number theorem, notably that $\pi(n)/(n/\log n) \to 1$ as $n \to \infty$

RH = Riemann('s) hypothesis that

$\zeta(s) = \Sigma n^{-s} = \Pi(1-p^{-s})^{-1}$

in which n runs through the natural numbers 1, 2, 3, ... and p through the primes 2, 3, 5, ... cannot be equal to zero for nonreal s other than of the form $s = 1/2 + iy$ with $i = \sqrt{-1}$ and y real.

Closing remarks

Publishing a book is a venture in which persons other than the author take part, and it is usual for the author to acknowledge their contributions where the list of those mentioned does not do so. My chief thanks for this edition are due to Thomas Wolf for typesetting it. He also typeset his translation in the German edition.

I am additionally indebted to James Flagg, who informed me that I had in effect proved Riemann's hypothesis in Appendix 7, and supplied me with the necessary works in analytic number theory to make the fact apparent.

My original purpose in Appendix 7 was to make a simpler formula than Riemann's for approximating the prime count at a given n. The formula I produced there is not entirely satisfactory because the correction factor t is not quite constant. It took me five more years to remedy this with the formula

$$S_d(n) = \sum_{k=1}^{[n^{1/2}]-1} \frac{2k+d}{\log(k+1)^2} + \frac{<n^{1/2}>(2[n^{1/2}]+d)}{\log([n^{1/2}]+1)^2} + \frac{1}{2}$$

where $[x]$ is the integer part and $<x>$the fraction part of x. The correction factor d is constant and can be set for any particular large n and will still give acceptably accurate answers for smaller n.

In the table that follows, my specification $j=11$ records how high we have taken k in Riemann's explicit formula (Appendix 7, formula (0)) to calculate the result in question, and $d = 0.06886$ is set to agree with Riemann's answer (using this j) for $n = 10^{12}$.

n	$\pi(n)$	$S_d(n)$	difference	$R_j(n)$	difference
50 000	5 133	5 133.691	+1	5 133.401	0
100 000	9 592	9 587.759	–4	9 587.434	–5
150 000	13 848	13 844.326	–4	13 843.985	–4
200 000	17 984	17 981.877	–2	17 981.534	–2
250 000	22 044	22 035.144	–9	22 034.818	–9
300 000	25 997	26 023.687	+27	26 023.336	+26
350 000	29 977	29 959.891	–17	29 959.550	–17
400 000	33 860	33 852.387	–8	33 852.051	–8
450 000	37 706	37 707.487	+1	37 707.114	+1
500 000	41 538	41 529.822	–8	41 529.508	–8
550 000	45 322	45 323.293	+1	45 322.978	+1
600 000	49 098	49 090.841	–7	49 090.533	–7
650 000	52 831	52 834.941	+4	52 834.644	+4
700 000	56 543	56 557.665	+15	56 557.737	+15
750 000	60 238	60 260.737	+23	60 260.465	+23
800 000	63 951	63 945.694	–5	63 945.415	–6
850 000	67 617	67 613.777	–3	67 613.519	–3
900 000	71 274	71 266.167	–8	71 265.905	–8
950 000	74 907	74 903.826	–3	74 903.572	–3
1 000 000	78 498	78 527.635	+30	78 527.402	+29
1 050 000	82 134	82 138.422	+4	82 138.184	+4
1 100 000	85 714	85 736.853	+23	85 736.626	+23
1 150 000	89 302	89 323.591	+22	89 323.367	+21
1 200 000	92 938	92 899.203	–39	92 898.986	–39
1 250 000	96 469	96 464.208	–5	96 464.011	–5
1 300 000	100 021	100 019.121	–2	100 018.925	–2
1 350 000	103 544	103 564.355	+20	103 564.170	+20
1 400 000	107 126	107 100.335	–26	107 100.153	–26
1 450 000	110 630	110 627.424	–3	110 627.250	–3
1 500 000	114 155	114 145.981	–9	114 145.811	–9
3 000 000	216 816	216 816.291	0	216 816.326	0
10^8	5 761 455	5 761 548.223	+93	5 761 551.872	+97
10^{12}	37 607 912 018	37 607 910 541.5	–1576	37 607 910 541.1	–1577

Table of comparisons of $\pi(n)$ with Spencer-Brown's $S_d(n)$ and the Riemann/Spencer-Brown $R_j(n)$ for small round values of n, using $d = 0.06886$ and $j = 11$. No estimate is wrong by more than about one tenth of one percent, and the last estimate is out by a factor of only 4.19×10^{-8}. Riemann's estimate is so complicated that we cannot predict exactly what will happen to it when n gets very large, but $S_d(n)$ is so simple that we can predict it will get better and better as n gets larger and larger.