THE DISTRIBUTION OF PRIME NUMBERS

INTRODUCTION

1. The positive integers other than 1 may be divided into two classes, prime numbers (such as 2, 3, 5, 7) which do not admit of resolution into smaller factors, and composite numbers (such as 4, 6, 8, 9) which do. The prime numbers derive their peculiar importance from the 'fundamental theorem of arithmetic' that a composite number can be expressed in one and only one way as a product of prime factors. A problem which presents itself at the very threshold of mathematics is the question of the distribution of the primes among the integers. Although the series of prime numbers exhibits great irregularities of detail, the general distribution is found to possess certain features of regularity which can be formulated in precise terms and made the subject of mathematical investigation.

We shall denote by $\pi(x)$ the number of primes not exceeding x; our problem then resolves itself into a study of the function $\pi(x)$. If we examine a table of prime numbers, we observe at once that, however extensive the table may be, the primes show no signs of coming to an end altogether, though they do appear to become on the average more widely spaced in the higher parts of the table. These observations suggest two theorems which may be taken as the starting-point of our subject. Stated in terms of $\pi(x)$, these are the theorems that $\pi(x)$ tends to infinity, and $\pi(x)/x$ to zero, as x tends to infinity.

2. The first theorem—that there exists an infinite number of primes—was proved by Euclid (*Elements*, Book 9, Prop. 20). In essentials his proof is as follows. Let P be a product of any finite set of primes, and let $Q = P + 1$. The integers P and Q can have no prime factor in common, since such a factor would divide $Q - P = 1$, which is impossible. But Q (being greater than 1) must be divisible by some prime. Hence there exists at least one prime distinct from those occurring in P. If there were only

a finite number of primes altogether, we could take P to be the product of all primes, and a contradiction would result. The argument really gives a little more. It shows that, if p_n is the nth prime (so that $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ...), the integer $Q_n = p_1 p_2 \dots p_n + 1$ is divisible by some p_m with $m > n$, so that $p_{n+1} \leqslant p_m \leqslant Q_n$; from which we may infer, by induction, that

$$p_n < 2^{2^n}. \tag{1}$$

3. In 1737 Euler proved the existence of an infinity of primes by a new method, which shows moreover that

$$\text{the series } \sum_{n=1}^{\infty} \frac{1}{p_n} \text{ is divergent.} \tag{2}$$

Euler's work is based on the idea of using an identity in which the primes appear on one side but not on the other. Stated formally his identity is

$$\sum_{n=1}^{\infty} n^{-s} = \prod_p (1 + p^{-s} + p^{-2s} + \dots) = \prod_p (1 - p^{-s})^{-1}, \tag{3}$$

where the products are over all primes p. Euler's contribution to the subject is of fundamental importance; for his identity, which may be regarded as an analytical equivalent of the fundamental theorem of arithmetic, forms the basis of nearly all subsequent work.

The theorems (1) and (2) resemble one another in that they each add something (though in different ways) to the statement that the number of primes is infinite.

4. The question of the diminishing frequency of primes was the subject of much speculation before any definite results emerged. The problem assumed a much more precise form with the publication by Legendre in 1808 (after a less definite statement in 1798) of a remarkable empirical formula for the approximate representation of $\pi(x)$. Legendre asserted that, for large values of x, $\pi(x)$ is approximately equal to

$$\frac{x}{\log x - B}, \tag{4}$$

where $\log x$ is the natural (Napierian) logarithm of x and B a

certain numerical constant[1]—a theorem described by Abel (in a letter written in 1823) as the 'most remarkable in the whole of mathematics'. A similar, though not identical, formula was proposed independently by Gauss. Gauss's method, which consisted in counting the primes in blocks of a thousand consecutive integers, suggested the function $1/\log x$ as an approximation to the average density of distribution ('number of primes per unit interval') in the neighbourhood of a large number x, and thus

$$\int_2^x \frac{du}{\log u} \tag{5}$$

as an approximation to $\pi(x)$. Gauss's observations were communicated to Encke in 1849, and first published in 1863; but they appear to have commenced as early as 1791 when Gauss was fourteen years old.[2] In the interval the relevance of the function (5) was recognised independently by other writers.[3] For convenience of notation it is usual to replace this function by the 'logarithmic integral'

$$\operatorname{li} x = \lim_{\eta \to +0} \left(\int_0^{1-\eta} + \int_{1+\eta}^x \right) \frac{du}{\log u},$$

from which it differs only by the constant $\operatorname{li} 2 = 1{\cdot}04\ldots$.

The precise degree of approximation claimed by Gauss and Legendre for their empirical formulae outside the range of the tables used in their construction is not made very explicit by either author, but we may take it that they intended to imply at any rate the 'asymptotic equivalence' of $\pi(x)$ and the approximating function $f(x)$, that is to say that $\pi(x)/f(x)$ tends to the limit 1 as x tends to infinity. The two theorems which thus arise, corresponding to the two forms of $f(x)$, are easily shown to be equivalent to one another and to the simpler relation

$$\frac{\pi(x)}{\dfrac{x}{\log x}} \to 1 \text{ as } x \to \infty; \tag{6}$$

but the distinction between (4) and (5), and the value of B in (4),

[1] Legendre **1a**, 19; **1b**, 394; **2**, ii, 65. The references in heavy type are to the bibliography at the end of the tract.

[2] Gauss **1**, ii, 444–447; $\mathbf{x_1}$, 11.

[3] Dirichlet, *Werke*, i, 372, footnote **; Chebyshev **1**, **2**; Hargreave **1**, **2**.

become important if we enquire more closely into the order of magnitude of the 'error' $\pi(x) - f(x)$. The proposition (6), which is now known as the 'prime number theorem', is the central theorem in the theory of the distribution of primes. The problem of deciding its truth or falsehood engaged the attention of mathematicians for about a hundred years.

5. The first theoretical results connecting $\pi(x)$ with $x/\log x$ are due to Chebyshev. In 1848 he showed (among other things) that, if the ratio on the left of (6) tends to a limit at all, the limit must be 1; and in 1850 that this ratio lies between two positive constants a and A for all sufficiently large values of x, so that the function $x/\log x$ does at any rate represent the true order of magnitude of $\pi(x)$. These results constituted an advance of the first importance, but (as Chebyshev himself was well aware) they failed to establish the essential point, namely the existence of $\lim \pi(x)/(x/\log x)$. And, although the numerical bounds (a, A) obtained by Chebyshev were successively narrowed by later writers (particularly Sylvester), it came to be recognised in due course that the methods employed by these authors were not likely to lead to a final solution of the problem.

6. The new ideas which were to supply the key to the solution were introduced by Riemann in 1859, in a memoir which has become famous, not only for its bearing on the theory of primes, but also for its influence on the development of the general theory of functions. Euler's identity had been used by Euler himself with a fixed value of s $(s = 1)$, and by Chebyshev with s as a real variable. Riemann now introduced the idea of treating s as a complex variable and studying the series on the left of (3) by the methods of the theory of analytic functions. This series converges only in a restricted portion of the plane of the complex variable s, but defines by continuation a single-valued analytic function regular at all finite points except for a simple pole at $s = 1$. This function is called the 'zeta-function of Riemann', after the notation $\zeta(s)$ adopted by its author.

Although Riemann is not primarily concerned with approximations to $\pi(x)$, his analysis shows clearly that this function is intimately bound up with the properties of $\zeta(s)$, and in par-

ticular with the distribution of its zeros in the s-plane. Riemann enunciated a number of important theorems concerning the zeta-function, together with a remarkable identity connecting $\pi(x)$ with its zeros, but he gave in most cases only insufficient indications of proofs. The problems raised by Riemann's memoir inspired in due course the fundamental researches of Hadamard in the theory of integral functions, the results of which at last removed some of the obstacles which for more than thirty years had barred the way to rigorous proofs of Riemann's theorems. The proofs sketched by Riemann were completed (in essentials), in part by Hadamard himself in 1893, and in part by von Mangoldt in 1894.

7. The discoveries of Hadamard prepared the way for rapid advances in the theory of the distribution of primes. The prime number theorem was proved in 1896 by Hadamard himself and by de la Vallée Poussin, independently and almost simultaneously. Of the two proofs Hadamard's is the simpler, but de la Vallée Poussin (in another paper published in 1899) studied in great detail the question of closeness of approximation. His results prove conclusively (what had been foreshadowed by Chebyshev) that, for all sufficiently large values of x, $\pi(x)$ is represented more accurately by $\operatorname{li} x$ than by the function (4) (no matter what value is assigned to the constant B), and that the most favourable value of B in (4) is 1. This conflicts with Legendre's original suggestion 1·08366 for B, but this value (based on tables extending only as far as $x = 400\,000$) had long been recognised as having little more than historical interest.

The theory can now be presented in a greatly simplified form, and de la Vallée Poussin's theorems can (if desired) be proved without recourse to the theory of integral functions. This is due almost entirely to the work of Landau. The results themselves underwent no substantial change until 1921, when they were improved by Littlewood; but Littlewood's refinements lie much deeper and the proofs involve very elaborate analysis.

8. The solution just outlined may be held to be unsatisfactory in that it introduces ideas very remote from the original problem, and it is natural to ask for a proof of the prime number theorem

not depending on the theory of functions of a complex variable. To this we must reply that at present no such proof is known. We can indeed go further and say that it seems unlikely that a genuinely 'real variable' proof will be discovered, at any rate so long as the theory is founded on Euler's identity. For every known proof of the prime number theorem is based on a certain property of the complex zeros of $\zeta(s)$, and this conversely is a simple consequence of the prime number theorem itself. It seems clear therefore that this property must be used (explicitly or implicitly) in any proof based on $\zeta(s)$, and it is not easy to see how this is to be done if we take account only of real values of s.

9. There is one important respect in which the theory is still very far from complete. Riemann conjectured (without any suggestion of proof) that the complex zeros of $\zeta(s)$, which (as Riemann proved) are confined to a certain infinite strip of the s-plane and lie symmetrically about the central line of this strip, are all situated on this central line. But this assertion, the now famous 'Riemann hypothesis', has never been proved or disproved, though the available evidence, both theoretical and numerical, seems to point in its favour. The truth of the Riemann hypothesis would entail considerable improvements of the theorems of de la Vallée Poussin and Littlewood on the order of magnitude of $\pi(x) - \operatorname{li} x$, but the true order cannot be decided so long as the truth of the hypothesis remains in doubt.

10. The relationship between $\pi(x)$ and $\operatorname{li} x$ is illustrated by the table on the opposite page (p. 7).[1] It will be noted at once that, for each value of x shown, $\pi(x) < \operatorname{li} x$. Until comparatively recently this inequality was believed to hold generally, and there were theoretical as well as numerical grounds for this belief; for the relation between $\pi(x)$ and $\zeta(s)$ associates $\pi(x)$ not directly with $\operatorname{li} x$, but with a more complicated expression of which the

[1] The last two entries lie outside the range of existing tables of primes, which stop a little beyond 100 000 000 [C. L. Baker and F. J. Gruenberger, *The first six million prime numbers*, The RAND Corporation, Santa Monica (published by The Microcard Foundation, Madison, Wisconsin, 1959)]. But $\pi(x)$ has been calculated for these values of x without actual enumeration of the primes; see D. H. Lehmer, *Illinois J. Math.*, 3 (1959), 381–388. The values of $\operatorname{li} x$ are given to the nearest integer. See also J. Glaisher **1**, 28–38; **2**, 66–103; D. N. Lehmer **1**, xiii–xvi; Phragmén **2**, 199–200 (footnote). [Footnote and table revised 1961.]

leading terms are $\operatorname{li} x - \frac{1}{2} \operatorname{li} x^{\frac{1}{2}}$. It was proved, however, by Littlewood (in 1914) that if we go far enough we shall eventually reach

x	$\pi(x)$	$\operatorname{li} x$	$\pi(x)/\operatorname{li} x$
1 000	168	178	0·94...
10 000	1 229	1 246	0·98...
50 000	5 133	5 167	0·993...
100 000	9 592	9 630	0·996...
500 000	41 538	41 606	0·9983...
1 000 000	78 498	78 628	0·9983...
2 000 000	148 933	149 055	0·9991...
5 000 000	348 513	348 638	0·9996...
10 000 000	664 579	664 918	0·9994...
20 000 000	1 270 607	1 270 905	0·9997...
90 000 000	5 216 954	5 217 810	0·99983...
100 000 000	5 761 455	5 762 209	0·99986...
1 000 000 000	50 847 534	50 849 235	0·99996...
10 000 000 000	455 052 512	455 055 614	0·999993...

a value of x for which $\pi(x) > \operatorname{li} x$, and that such values will recur infinitely often. Littlewood's theorem is a pure 'existence theorem', and we still know no numerical value of x for which $\pi(x) > \operatorname{li} x$. It is probable that the first such value lies far beyond the range of the above table.

There is a similar phenomenon in connection with the distribution of the odd primes between the two arithmetical progressions $4n+1$ and $4n+3$. If $\pi^{(1)}(x)$ and $\pi^{(3)}(x)$ denote respectively the number of primes of these two forms which do not exceed x, then $\pi^{(1)}(x)/\pi^{(3)}(x)$ tends to the limit 1 as x tends to infinity. (This theorem is of the same 'depth' as the prime number theorem, and its proof depends on the theory of functions of a complex variable.) Thus, to a first approximation, the odd primes are evenly distributed between the two progressions. But the tables show a definite preponderance of primes of the form $4n+3$, and until 1914 all the available evidence pointed to the conclusion that (except for a short range at the beginning) $\pi^{(1)}(x) < \pi^{(3)}(x)$.[1] But Littlewood's method shows that $\pi^{(1)}(x) > \pi^{(3)}(x)$ for arbitrarily large values of x, though again it provides no numerical solution of this inequality.

[1] J. W. L. Glaisher 1. An earlier table by Scherk (*Journal für Math.*, 10 (1833), 208) is very inaccurate.

11. The present tract is devoted to a systematic study of the asymptotic relations for $\pi(x)$ discussed in outline in the foregoing sections. The theory of the Riemann zeta-function will be developed only so far as it is required for applications to $\pi(x)$. The more advanced theory of $\zeta(s)$ forms the subject of a companion volume by E. C. Titchmarsh.[1]

There are many questions relating to the distribution of primes which we are unable to discuss in this tract, either through lack of space or because the problems are as yet unsolved. To the former category belongs the general theory of the distribution of the primes among the various arithmetical progressions of given difference k, a theory associated principally with the names of Dirichlet and de la Vallée Poussin.[2] To the second category belong nearly all questions relating to the finer structure of the series of primes. The prime number theorem shows that the average interval $p_{n+1}-p_n$ between (large) consecutive primes is about $\log p_n$, but there may be wide deviations in either direction from this average. There are strong indications on the one hand that the interval reduces infinitely often to the value 2, so that there exists an infinity of pairs of primes differing only by 2 (such as 17, 19 or 10 006 427, 10 006 429); but this has not been proved. The opposite problem, that of abnormally large values of $p_{n+1}-p_n$, is also unsolved, and such indications as do exist are of a negative character. Thus, on the Riemann hypothesis, we infer easily from the results of this tract that, if ϑ is any fixed number greater than $\frac{1}{2}$, the interval is never as large as p_n^{ϑ} except possibly in a finite number of instances, and it has been conjectured that the stronger assertion with $\vartheta=\frac{1}{2}$ is also true; but the most that has actually been proved is the corresponding statement with $\vartheta > 1-(33000)^{-1}$.[3]

[1] T in the bibliography at the end of this tract.

[2] For an account of this theory (and for a full treatment of the subject as a whole) we refer to the two well-known books of Landau (H and V in the bibliography).

[3] For the last result see Hoheisel 1. For an account of the delicate problems just referred to, and of related problems, see V; BC, 805–810, and the references there given; Hardy and Littlewood, 4, 5; Hardy and Littlewood, *Proc. London Math. Soc.* (2), 28 (1928), 518 (footnote); Schnirelmann 1; Landau 10.

CHAPTER I

ELEMENTARY THEOREMS

1. In this chapter we confine ourselves to theorems which can be proved without the use of the theory of functions of a complex variable. The main results are superseded by those of later chapters, but the elementary arguments are of great interest on account of their simplicity and directness.

We denote primes generally by p, and the nth prime by p_n. We denote by $\pi(x)$ the number of primes not exceeding x, where x is a positive number (not necessarily integral). We shall use the notations

$$\sum_{n \leqslant x} f(n), \quad \sum_{p > x} f(p), \quad \prod_{p} f(p), \text{ etc.}$$

(with various modifications and extensions which will be explained by the context) to indicate sums or products over all positive integers n, or all primes p, within the specified ranges; in the third example, where no range is indicated, it is understood that all primes are included. The order of the terms or factors (when relevant) is that which corresponds to increasing n or p. We adopt the general convention that an 'empty' sum (i.e. a sum containing no terms) is to have the value 0, and an 'empty' product the value 1. As examples we have (for $x > 0$)

$$[x] = \sum_{n \leqslant x} 1, \quad \pi(x) = \sum_{p \leqslant x} 1, \quad [x]! = \prod_{n \leqslant x} n,$$

where $[u]$ denotes (for any real u) the 'integral part' of u (i.e. the integer m defined by $m \leqslant u < m + 1$). We sometimes write

$$\sum_{n=1}^{x} \text{ for } \sum_{n=1}^{[x]}.$$

We use the symbols O, o, and $\sim$ (the sign of 'asymptotic equality') in the senses which are now classical. Thus

$$f(x) = O(x), \quad f(x) = o(x), \quad f(x) \sim x,$$

(as $x \to \infty$) mean respectively '$|f(x)|/x$ is less than a constant K (i.e. a number independent of x) for all sufficiently large x', '$f(x)/x \to 0$ as $x \to \infty$', and '$f(x)/x \to 1$ as $x \to \infty$'.

The notations $m \mid n$ and $m \nmid n$ mean 'm divides n' and 'm does not divide n' respectively.

2. Theorem 1. *The series and product*

$$\sum_p \frac{1}{p}, \qquad \prod_p \left(1-\frac{1}{p}\right)^{-1}$$

are divergent.[1]

Write

$$S(x) = \sum_{p \leqslant x} \frac{1}{p}, \quad P(x) = \prod_{p \leqslant x} \left(1-\frac{1}{p}\right)^{-1} \qquad (x > 2).$$

Since

$$1/(1-u) > (1-u^{m+1})/(1-u) = 1 + u + \ldots + u^m \qquad (0 < u < 1),$$

we have

$$P(x) > \prod_{p \leqslant x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \ldots + \frac{1}{p^m}\right),$$

where m is any positive integer. Now the product on the right, when multiplied out, is equal to $\Sigma\, 1/n$ summed over a certain set of positive integers n, and, if m is chosen so that $2^{m+1} > x$, this set will certainly include all integers from 1 to $[x]$. Hence

$$(1) \qquad P(x) > \sum_1^{[x]} \frac{1}{n} > \int_1^{[x]+1} \frac{du}{u} > \log x.$$

Since $-\log(1-u) - u < \frac{1}{2}u^2/(1-u)$ for $0 < u < 1$ (from the series for $\log(1-u)$), we have

$$\log P(x) - S(x) < \sum_{p \leqslant x} \frac{p^{-2}}{2(1-p^{-1})} < \sum_{n=2}^{\infty} \frac{1}{2n(n-1)} = \tfrac{1}{2}.$$

Hence, by (1),

$$(2) \qquad S(x) > \log\log x - \tfrac{1}{2}.$$

The inequalities (1) and (2) evidently establish the theorem.

3. Theorem 1 shows incidentally that the number of primes is infinite, or that $\pi(x)$ tends to infinity with x. We next show that $\pi(x)/x$ tends to zero.

Theorem 2. $\pi(x) = o(x)$ *as* $x \to \infty$.

Denote by $N_r(x, h)$, where $x > 0$ and h is a positive integer, the number of positive integers n not exceeding x which are divisible by h but not by any of the first r primes $p_1, \ldots, p_r$, $N_0(x, h)$ being simply the number of $n \leqslant x$ divisible by h. Then,

[1] Euler 1, Theorema 19; 2, § 279.