ON THE EXACT NUMBER OF PRIMES LESS THAN A GIVEN LIMIT

BY
D. H. LEHMER

The problem of counting the exact number of primes $\leqq x$, without actually listing them all, dates from Legendre [1] who observed that the number of primes p for which $x^{1/2} < p \leqq x$ is one less than

$$x - \sum_i [x/p_i] + \sum_{i<j} [x/p_i p_j] - \cdots ,$$

where $[z]$ denotes, as usual, the greatest integer $\leqq z$, and the p_i range over all the primes less than or equal to $x^{1/2}$. Since then, a large number of writers [2] have suggested variants and improvements of this result. Foremost among these was the astronomer Meissel [3] whose formula (5) is derived below. This formula, and Meissel's obscure derivation of it, is to be found in a number of textbooks in number theory. It is of practical value to our problem because in its "Legendre's sum" the primes extend only as far as $x^{1/3}$. Meissel used his formula to evaluate $\pi(x)$, the number of primes $\leqq x$ for a number of large values of x including

$$\pi(10^7) = 664\,579, \qquad \pi(10^8) = 5\,761\,455, \qquad \pi(10^9) = 50\,847\,478.$$

Others writers with "better" formulas than Legendre's or Meissel's have been content to advocate rather than utilize their results. At any rate, until now no one has made such calculations beyond $x = 10^7$, except N. P. Bertelsen [4] who confirmed Meissel's corrected value of $\pi(10^8)$ and computed $\pi(2 \cdot 10^7)$ and $\pi(9 \cdot 10^7)$. Meissel's calculation of $\pi(10^9)$, made sometime between 1871 and 1885, must be regarded as one of the outstanding single calculations of the 19th century, even though his value is slightly in error. Because of the recent interest in such functions as $\pi(x) - \mathrm{li}(x)$, the writer has been considering the problem of extending Meissel's method so as to reduce the range of the primes in the Legendre sum still further. The fact that we now have high speed computers to do our actual calculations does not relieve us of the responsibility of counting the minutes as Meissel must have counted his weeks. Some preliminary work on the SWAC in 1956 indicated that for very fast machines there is a decided lack of balance in Meissel's formula, most of the time being spent on its Legendre sum.

Notation and general formula

Let m_a denote the product

$$m_a = p_1 p_2 \cdots p_a$$

of the first a primes. We consider the general Legendre sum

$$\phi(x, a) = \sum \mu(\delta)[x/\delta],$$

Received April 1, 1958.

where δ ranges over all the divisors of m_a, and $\mu(\delta)$, the Möbius function, has, in this case, the value $+1$ or -1 according as δ is the product of an even or an odd number of primes. By the well-known combinatorial principle, called by Sylvester the Principle of Crossclassification, $\phi(x, a)$ is the number of numbers $\leqq x$ divisible by none of the first a primes.

We denote by $P_k(x, a)$ the number of products $\leqq x$ of k primes each greater than p_a. We take $P_0(x, a) = 1$ and note that $P_k(x, a) = 0$ for all $k \geqq r$ where $x < p_{a+1}^r$. Since each of the numbers enumerated by $\phi(x, a)$ is a product of a certain number of primes $> p_a$, we have at once the general formula

$$\phi(x, a) = \sum_{k=0}^{r-1} P_k(x, a). \tag{1}$$

Because we wish to find $\pi(x)$, our interest in this formula stems from the simple fact that

$$P_1(x, a) = \pi(x) - a, \tag{2}$$

and our formula becomes effective as soon as we provide adequate means of computing the other P's as well as an independent practical method of evaluating $\phi(x, a)$. With x given, our choice of a and hence the number of non-zero P's will depend on the economics of these separate parts of the calculation.

The functions $P_k(x, a)$

These functions may be evaluated from a knowledge of a more or less limited list of primes or of the function $\pi(y)$. Thus for $k = 2$ we have

$$\begin{aligned} P_2(x, a) &= \sum_{p_1 p_j \leqq x} 1 = \sum_{p_a < p_i \leqq x/p_i} \sum_{p_i \leqq p_j \leqq x/p_i} 1 \\ &= \sum_{p_a < p_i \leqq x^{1/2}} \{\pi(x/p_i) - (i - 1)\}. \end{aligned}$$

If we denote the number of primes not exceeding the square root and cube root of x by

$$b = \pi(x^{1/2}) \quad \text{and} \quad c = \pi(x^{1/3}),$$

then we may write the above as follows:

$$P_2(x, a) = \sum_{a<i\leqq b} \pi(x/p_i) - \tfrac{1}{2}(b - a)(b + a - 1). \tag{3}$$

For $k = 3$ we have

$$\begin{aligned} P_3(x, a) &= \sum_{p_a < p_i \leqq x/p_i^2} \sum_{p_i \leqq p_j \leqq x/p_i p_j} \sum_{p_j \leqq p_k \leqq x/p_i p_j} 1 \\ &= \sum_{a<i\leqq c} \sum_{i \leqq j \leqq b_i} \{\pi(x/p_i p_j) - (j - 1)\}, \end{aligned} \tag{4}$$

where we have introduced

$$b_i = \pi\{(x/p_i)^{1/2}\}.$$

In using this formula for $P_3(x, a)$ it is supposed that a short table of $\pi(y)$ is stored in the high-speed memory of the machine for rapid consultation at very frequent intervals of time. This is not so in the formula for $P_2(x, a)$. Here the argument x/p_i may be expected, for a small, to reach high values beyond the storage capacity of the machine's fast memory. The procedure to use in this case will be discussed later.

To derive Meissel's formula we have only to set $a = c$. Then $P_3(x, c) = 0$, and the general formula (1), when solved for $\pi(x)$, becomes in view of (2) and (3)

$$\pi(x) = \phi(x, c) - 1 + c - P_2(x, c),$$

or

$$\pi(x) = \phi(x, c) + \tfrac{1}{2}(b + c - 2)(b - c + 1) - \textstyle\sum_{c<i\leqq b} \pi(x/p_i), \tag{5}$$

which is Meissel's formula.

Another special case arises when we put

$$a = \pi(x^{1/4}).$$

In this case $P_4(x, a) = 0$, and the general formula becomes in view of (2) and (4)

$$\begin{aligned}\pi(x) = \phi(x, a) &+ \tfrac{1}{2}(b + a - 2)(b - a + 1) - \textstyle\sum_{a<i\leqq b} \pi(x/p_i) \\ &- \textstyle\sum_{a<i\leqq c} \sum_{i\leqq j\leqq b_i} \{\pi(x/p_i p_j) - (j - 1)\}.\end{aligned} \tag{6}$$

Finally if we set $a = b$, we obtain

$$\pi(x) = \phi(x, b) - 1 + b,$$

which is Legendre's formula.

Evaluation of $\phi(x, a)$

The number $\phi(u, \nu)$ of numbers $\leqq u$ not divisible by the first ν primes satisfies the recurrence

$$\phi(u, \nu) = \phi(u, \nu - 1) - \phi(u/p_\nu, \nu - 1). \tag{7}$$

In fact the numbers enumerated by $\phi(u, \nu - 1)$ are of two sorts: those not divisible by p_ν, of which there are $\phi(u, \nu)$, and those divisible by p_ν of which there are $\phi(u/p_\nu, \nu - 1)$.

If we use (7) repeatedly starting with $u = x$, $\nu = a$, we get an aggregate of terms of the form

$$(-1)^L \phi(x/(p_{a_1} p_{a_2} \cdots p_{a_L}), \lambda),$$

where

$$a_1 < a_2 < \cdots < a_L \leqq \lambda \leqq a.$$

Such a term may be said to be of level L and signature $(-1)^L$. There are three ways of disposing of this term. In the first place the first argument of ϕ may be small, in fact less than p_λ. In this case $\phi = 1$ by definition.

In the second place, the second argument of ϕ may be small, so that ϕ may be found by having the machine consult a stored table of $\phi(u, k)$ for fixed k.

This value may then be accumulated, according to its signature, as a contribution to the final value of $\phi(x, a)$. A good practical value of k is 5. Suppose

$$u = 2310q + r \qquad (0 \leqq r \leqq 2309).$$

Then

$$\phi(u, 5) = 480q + \phi(r, 5), \tag{8}$$

where the last term is tabulated. Only 1155 entries in this table are needed since

$$\phi(2s, 5) = \phi(2s - 1, 5).$$

The table could be cut in half again by exploiting the functional equation

$$\phi(2309 - u, 5) = 480 - \phi(u, 5)$$

if space for storage is a problem. It is perhaps worth pointing out that the multiplication by 480 indicated in (8) may be postponed until the very end of the calculation by accumulating all the items q, according to their signature, and multiplying their accumulated value by 480 only once. This serves to speed up an already lengthy calculation.

Finally if neither argument of ϕ is sufficiently small, the recurrence (7) is resorted to; thus replacing λ by $\lambda - 1$ in our term and introducing the new term

$$(-1)^{L+1}\phi(x/(p_\lambda p_{a_1} \cdots p_{a_L}), \lambda - 1)$$

of level $L + 1$ at the cost of one division operation. This new term is the one to which the machine now gives its undivided attention.

Thus in carrying out the calculation of $\phi(x, a)$ in this methodical way, there is in storage, at any one time, only one term of level 0, one term of level 1, and so on up to one term of highest level L, the latter being treated in one of the three ways mentioned above, while the terms of lower level await future consideration. The number of terms in storage is quite modest, 10 levels being sufficient to treat an x as large as $5 \cdot 10^{12}$, assuming a table of $\phi(r, 5)$ is also stored.

The calculation of $\phi(x, a)$, if written out in full for a large x and a modest value of a, would be a highly ramified structure reminiscent of a shower or cascade of particles produced by a single cosmic ray. Each application of (7) gives a bifurcation into a pair of terms, one being of the next level. This process continues until the term is absorbed by the boundary conditions $u < p_\nu$ or $\nu = 5$. To give an idea of the extent of this proliferation, for the case of $\phi(10^{10}, 65)$ there were precisely 2 818 344 consultations of the table of $\phi(u, 5)$ and 401 962 cases in which $u < p_\nu$, giving in all 3 220 306 ends of branches, yet there were no terms of level 8 or more.

This is a good example of how one can substitute time for space with a high-speed computer. For checking purposes one can use the fact that

$\phi(x, a)$ is approximately equal to

$$x \prod_{i \leqq a} (1 - p_i^{-1}).$$

Tables of this product have been given by Legendre [5] and Glaisher [6] for $a \leqq 1229$ and 10 000 respectively. This approximation is often fairly good.

It was in the $\phi(x, a)$ part of the calculation of $\pi(10^9)$ by means of (5) that Meissel made his mistake. His value of $\phi(10^9, 168)$ is 57 short of its true value 81 515 102. A recalculation of $\pi(10^9)$ by means of (6) confirms this fact. There is also a minor error of a unit in his value of $\sum \pi(x/p_i)$.

The calculation of $P_2(x, a)$

In calculating $P_2(x, a)$ the sum

$$\sum_{a < i \leqq b} \pi(x/p_i) \tag{9}$$

is best found by accumulating tallies of the number of primes between x/p_i and x/p_{i-1}. To this effect we first construct a large number of binary digits in the number

$$.11101101101001100101101 00 \cdots \tag{10}$$

whose k^{th} digit is 1 or 0 according as $2k + 1$ is a prime or not. This can be done by a straightforward Eratosthenes sieve process running over a memory-ful of binary digits at one time. The characteristic number may then be stored on magnetic tape in lieu of a list of primes. Four large tapes cover the primes less than 50 million. In evaluating the sum (9) the p's are run backward from p_b to p_a, and the long binary number (10) is fed through the arithmetic unit which counts the binary 1's. At the appropriate moments the subtotal is accumulated until the final sum is reached. In programming the change from one prime p to the next smaller one, a small copy of the characteristic number (10) is fed backwards through the arithmetic unit.

A similar procedure applies to the calculation of $P_3(x)$. In this case, as mentioned above, a small table of $\pi(x)$ is stored for accumulation as called for in (4).

The results shown in Table 1 illustrate the general formula (1).

Using (1) and (2) we derive at once from the values in Table 1 the values of $\pi(x)$ shown in Table 2. These are compared with the approximations

$$\mathrm{li}(x) = \int_2^x dt/\log t \quad \text{and} \quad R(x) = \sum_{n=1}^{\infty} n^{-1}\mu(n)\, \mathrm{li}(x^{1/n}),$$

the latter being suggested by Riemann [7].

It will be noted that in the cases of $x = 90\,000\,000$ and $x = 1\,000\,000\,000$ where there are two sets of data, both sets lead to exactly the same value of $\pi(x)$.

The values of $\pi(x)$ for $x = 20\,000\,000$ and $90\,000\,000$ are in agreement with those of Bertelsen.

TABLE 1

x	a	$\phi(x, a)$	$P_2(x, a)$	$P_3(x, a)$
15 485 864	53	1 568 715	568 767	0
20 000 000	58	1 988 057	717 507	0
25 000 000	61	2 458 751	892 884	0
32 452 845	66	3 140 783	1 140 848	0
33 000 000	66	3 193 726	1 162 124	0
37 000 000	67	3 569 832	1 308 275	0
40 000 000	68	3 847 872	1 414 285	0
90 000 000	25	10 826 326	5 127 120	482 276
90 000 000	86	8 270 815	3 053 946	0
100 000 000	90	9 110 819	3 349 453	0
999 000 000	40	107 108 994	51 195 835	5 113 621
1 000 000 000	40	107 216 231	51 248 370	5 120 366
1 000 000 000	168	81 515 102	30 667 735	0
10 000 000 000	65	964 916 391	463 026 862	46 837 081

TABLE 2

x	$\pi(x)$	$\mathrm{li}(x) - \pi(x)$	$R(x) - \pi(x)$
15 485 864	1 000 000	411	109
20 000 000	1 270 607	298	−37
25 000 000	1 565 927	369	2
32 452 845	2 000 000	354	−56
33 000 000	2 031 667	307	−105
37 000 000	2 261 623	630	194
40 000 000	2 433 654	362	−85
90 000 000	5 216 954	399	227
100 000 000	5 761 455	755	97
999 000 000	50 799 577	1389	−387
1 000 000 000	50 847 534	1701	−79
10 000 000 000	455 052 512	3102	−1829

If we take the difference

$$\pi(1\,000\,000\,000) - \pi(999\,000\,000),$$

we find 47 957 as the number of primes in the thousandth million. This agrees with the result obtained by F. Gruenberger [8] who has made a special study of this million. This is another confirmation of the fact that Meissel's often quoted value 50 847 478 for $\pi(10^9)$ is too low by precisely 56.

It will be noted that for the above values of x the function $R(x)$ is a much better approximation than $\mathrm{li}(x)$. This is not always the case, however. According to a theorem of Littlewood, the difference $\mathrm{li}(x) - \pi(x)$ changes sign infinitely often. Since $R(x)$ is always less than $\mathrm{li}(x)$, it must be a worse approximation infinitely often. Recent results of Skewes [9] do not give us much encouragement in looking for such an occurrence with the methods and equipment available to us at this time.

The general sum over primes

We conclude with a few remarks on the extension of the preceding formulas to the problem of summing a function over primes, for example, finding the sums of k^{th} powers of primes $\leqq x$.

Let $f(n)$ be any numerical function (which up to now has been identically equal to 1), and let $Q(x, k)$ be defined for $x > 0$ and k a positive integer by

$$Q(x, k) = \sum_{nk \leqq x} f(nk)$$

summed over all multiples of k not exceeding x. The functions $\phi(x, a)$ and $P_k(x, a)$ are defined by

$$\phi(x, a) = \sum_{n \leqq x, (n, m_a)=1} f(n) \quad (m_a = p_1 p_2 \cdots p_a),$$

$$P_0(x, a) = f(1), \qquad P_k(x, a) = \sum_{q_k \leqq x} f(q_k),$$

where q_k is a product of k primes each greater than p_a. Then, as a generalization of (1), we have

$$\phi(x, a) = \sum_{\delta \mid m_a} \mu(\delta) Q(x, \delta) = \sum_{k=0}^{r-1} P_k(x, a). \tag{11}$$

To prove the first equality we use the well-known property of Möbius' function:

$$\sum_{\delta \mid N} \mu(\delta) = [1/N], \tag{12}$$

and write

$$\sum_{\delta \mid m_a} \mu(\delta) Q(x, \delta) = \sum_{\delta \mid m_a} \mu(\delta) \sum_{n\delta \leqq x} f(n\delta)$$
$$= \sum_{m \leqq x} f(m) \sum_{n\delta = m, \delta \mid m_a} \mu(\delta).$$

By (12) the inner sum is seen to vanish except when m and m_a are relatively prime, in which case the coefficient of $f(m)$ is unity. This gives us $\phi(x, a)$ by definition. The second equality of (11) follows at once from the simple observation that each of the numbers $\leqq x$ and prime to m_a is uniquely the product of a certain number, k, of prime factors each greater than p_a. The relation (11) contains

$$P_1(x, a) = \sum_{p_a < p \leqq x} f(p),$$

and thus the calculation of $\sum f(p)$ over primes $p \leqq x$ can be made to depend upon $\phi(x, a)$ and the other P's, as before.

As to the evaluation of $\phi(x, a)$, a recurrence like (7) exists when f is "purely" multiplicative, that is,

$$f(m) f(n) = f(mn) \tag{13}$$

for every pair (m, n) of integers, as for example $f(n) = n^k$. In fact in this case

$$\phi(x, a) = \phi(x, a - 1) - f(p_a)\phi(x/p_a, a - 1),$$

which may be programmed as in the previous discussion.

If f is only multiplicative, that is, (13) holds in case m and n are coprime, then we have recurrence

$$\phi(x, a) + f(p_a)\phi(x/p_a, a) + f(p_a^2)\phi(x/p_a^2, a) + \cdots = \phi(x, a - 1).$$

In this case the programming is much more elaborate for large x.

The calculations outlined above were carried out at the University of California Computing Center on the Berkeley campus on an IBM 701 computer. Some of the results were obtained while the center was partially supported by the National Science Foundation.

References

1. A.-M. Legendre, *Théorie des nombres*, 2 éd., Paris, 1808, p. 420.
2. L. E. Dickson, *History of the theory of numbers*, Washington, 1919, vol. 1, pp. 429–435.
   A. Brauer, *On the exact number of primes below a given limit*, Amer. Math. Monthly, vol. 53 (1946), pp. 521–523.
3. E. D. F. Meissel, *Ueber die Bestimmung der Primzahlenmenge innerhalb gegebener Grenzen*, Math. Ann., vol. 2 (1870), pp. 636–642.
   ———, *Berechnung der Menge von Primzahlen, welche innerhalb der ersten Hundert Millionen natürlicher Zahlen vorkommen*, Math. Ann., vol. 3 (1871), pp. 523–525.
   ———, *Ueber Primzahlmengen*, Math. Ann., vol. 21 (1883), p. 304.
   ———, *Berechnung der Menge von Primzahlen, welche innerhalb der ersten Milliarde natürlicher Zahlen vorkommen*, Math. Ann., vol. 25 (1885), pp. 251–257.
4. J.-P. Gram, *Rapport sur quelques calculs entrepris par M. Bertelsen et concernant les nombres premiers*, Acta Math., vol. 17 (1893), pp. 301–314.
5. A.-M. Legendre, loc. cit., Table IX.
6. J. W. L. Glaisher, *Table of the values of* $\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{4}{5}\cdots\frac{x-1}{x}$, *the denominators being the series of prime numbers*, Messenger of Mathematics, vol. 28 (1898), pp. 2–15.
7. G. F. B. Riemann, *Über die Anzahl der Primzahlen unter einer gegeben Grösse*, Monatsberichte Akad. Wiss. Berlin, 1859, pp. 671–680.
8. C. L. Baker and F. J. Gruenberger, *Primes in the thousandth million*, Santa Monica, Rand Corporation, 1958, Publication 1245.
9. S. Skewes, *On the difference* $\pi(x) - \mathrm{li}\, x$ *(II)*, Proc. London Math. Soc. (3), vol. 5 (1955), pp. 48–70.

University of California
Berkeley, California