TECHNICAL NOTES AND SHORT PAPERS

# A Heuristic Asymptotic Formula Concerning the Distribution of Prime Numbers

By Paul T. Bateman and Roger A. Horn

Suppose $f_1, f_2, \cdots, f_k$ are polynomials in one variable with all coefficients integral and leading coefficients positive, their degrees being $h_1, h_2, \cdots, h_k$ respectively. Suppose each of these polynomials is irreducible over the field of rational numbers and no two of them differ by a constant factor. Let $Q(f_1, f_2, \cdots, f_k; N)$ denote the number of positive integers $n$ between 1 and $N$ inclusive such that $f_1(n), f_2(n), \cdots, f_k(n)$ are all primes. (We ignore the finitely many values of $n$ for which some $f_i(n)$ is negative.) Then heuristically we would expect to have for $N$ large

$$(1) \qquad Q(f_1, \cdots, f_k; N) \sim h_1^{-1}h_2^{-1} \cdots h_k^{-1} C(f_1, \cdots, f_k) \int_2^N (\log u)^{-k}\, du,$$

where

$$C(f_1, f_2, \cdots, f_k) = \prod_p \left\{\left(1 - \frac{1}{p}\right)^{-k}\left(1 - \frac{\omega(p)}{p}\right)\right\},$$

the product being taken over all primes and $\omega(p)$ being the number of solutions of the congruence

$$f_1(x)\, f_2(x) \cdots f_k(x) \equiv 0 \pmod{p}.$$

(If $\omega(p) = p$ for some prime $p$, then $C(f_1, f_2, \cdots, f_k) = 0$ and $Q(f_1, f_2, \cdots, f_k) \leqq h_1 + h_2 + \cdots + h_k$ for all $N$; we agree to exclude this trivial case.) Although it does not seem likely that (1) will be proved in the foreseeable future (aside from the known case of a single linear polynomial), a simple application of Atle Selberg's sieve method [1] does show that

$$(2) \quad Q(f_1, \cdots, f_k; N) \leqq 2^k k!\, C(f_1, \cdots, f_k) \int_2^N (\log u)^{-k}\, du + o(N(\log N)^{-k}).$$

The details of the argument were given by Bateman and Stemmler [2]. Thus we at least know an upper bound for $Q(f_1, f_2, \cdots, f_k; N)$ which is $2^k\, k!\, h_1h_2 \cdots h_k$ times the conjectured asymptotic value.

The heuristic argument in support of (1) may be put into various forms, but essentially amounts to the following. In some sense the chance that a large positive integer $m$ is prime is around $1/\log m$. Since $\log f_i(n)$ is around $h_i \log n$, the chance that $f_1(n), f_2(n), \cdots, f_k(n)$ are all primes would seem to be about

$$h_1^{-1}h_2^{-1} \cdots h_k^{-1} (\log n)^{-k}.$$

Received November 20, 1961. This work was supported by the National Science Foundation and the Office of Naval Research. The authors would like to thank Professor Atle Selberg for some helpful remarks.

But this ignores the fact that $f_1(n), f_2(n), \cdots, f_k(n)$ are not quite random integers. Thus for each prime $p$ we must apply a correction factor $r_p/s_p$, where $r_p$ is the chance that for random $n$ none of the integers $f_1(n), f_2(n), \cdots, f_k(n)$ is divisible by $p$ and $s_p$ is the chance that none of the integers in a random $k$-tuple is divisible by $p$. But clearly $r_p = 1 - \omega(p)/p$ and $s_p = (1 - 1/p)^k$. Thus the chance that $f_1(n), f_2(n), \cdots, f_k(n)$ are all primes for a large positive integer $n$ taken at random is about

$$h_1^{-1} h_2^{-1} \cdots h_k^{-1} C(f_1, \cdots, f_k) (\log n)^{-k}.$$

Hence we would expect $Q(f_1, f_2, \cdots, f_k ; N)$ to be about

$$h_1^{-1} h_2^{-1} \cdots h_k^{-1} C(f_1, \cdots, f_k) \sum_{n=2}^{N} (\log n)^{-k},$$

which is essentially the same as the approximation given in (1). This "derivation" of the heuristic formula (1) by the simple expedient of multiplying $p$-adic densities has the virtue that an analogous method gives results known to be correct in Waring's problem and in the theory of quadratic forms.

The convergence of the product defining $C(f_1, \cdots, f_k)$ may be proved as follows. Let $\omega_i(p)$ be the number of solutions of the congruence

$$f_i(x) \equiv 0 \pmod{p}.$$

If $i \neq j$, there exist polynomials $g_{ij}$ and $h_{ij}$ with integral coefficients and a positive integer $c_{ij}$ such that

$$g_{ij}(x) f_i(x) + h_{ij}(x) f_j(x) = c_{ij}$$

identically in $x$. Thus

$$\omega(p) = \omega_1(p) + \cdots + \omega_k(p)$$

for all sufficiently large primes $p$, namely for all primes $p$ such that

$$p \nmid \prod_{1 \leq i < j \leq k} c_{ij}.$$

For all but a finite number of primes $p$ we also know that $\omega_i(p)$ is the number of distinct prime ideals of norm $p$ in the algebraic number field generated by a zero of $f_i$. Hence by an elementary result on the distribution of prime ideals we have for $x$ large

$$\sum_{p \leq x} \frac{\omega_i(p)}{p} = \log \log x + A_i + o(1)$$

$$= \sum_{p \leq x} \frac{1}{p} + B_i + o(1),$$

where $A_i$ and $B_i$ are certain constants. Thus the series

$$\sum_{p} \frac{\omega_i(p) - 1}{p}$$

converges for $i = 1, \cdots, k$ and so

$$\sum_{p} \frac{\omega(p) - k}{p}$$

converges. Since

$$\left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{\omega(p)}{p}\right)=1+\frac{k-\omega(p)}{p}+\frac{B(p)}{p^2},$$

where $B(p)$ is a bounded function of $p$, it follows that the product for $C(f_1, \cdots, f_k)$ converges.

Numerical evidence has been given for many special cases of (1), particularly when the $f_i$ are all linear polynomials [3], [4]. The case of the single polynomial $n^2+1$ was discussed by Western [5] and other polynomials of the form $n^2+a$ were considered by Shanks [6]. The case of the pair of polynomials $n^2-2n+2$ and $n^2+2n+2$ was discussed by Shanks [7]. The case of the single polynomial $n^4+1$ was treated by Shanks [8]. In all cases the heuristic formula (1) fits the numerical data remarkably well.

In this note we give some numerical data for the case of the pair of polynomials $n$ and $n^2+n+1$. That is, we consider the number $P(N)$ of primes $p$ not exceeding $N$ such that $p^2+p+1$ is also a prime. The heuristic formula here is

$$P(N)\sim\frac{1}{2}C\int_2^N(\log u)^{-2}\,du, \tag{3}$$

where

$$C=\prod_p\left\{\left(1-\frac{1}{p}\right)^{-2}\left(1-\frac{2+\chi(p)}{p}\right)\right\},$$

where $\chi(p)$ is $-1$, 0, or $+1$ according as $p$ is congruent to $-1$, 0, or $+1$ modulo 3. As in every case where each of the polynomials $f_i$ has the property that a zero thereof generates a normal extension of the field of rational numbers with an abelian galois group, we can express $C$ in terms of an absolutely convergent infinite product. Specifically

$$C=L(1,\chi)^{-1}\prod_p\left\{1-\frac{p+2p\chi(p)-\chi(p)}{(p-1)^2(p-\chi(p))}\right\},$$

where

$$L(1,\chi)=\prod_p\left(1-\frac{\chi(p)}{p}\right)^{-1}=\sum_{n=1}^{\infty}\frac{\chi(n)}{n}=\frac{\pi}{3\sqrt{3}}.$$

The numerical value of $C$ is about 1.522.

The second-named author used the ILLIAC to prepare a list of the 776 primes of the form $p^2+p+1$ with $p$ a prime less than 113,000. (The program used was a straightforward one, and the running time was about 400 minutes.) The first 209 of these primes are listed by Bateman and Stemmler [2] who considered primes of the form $p^2+p+1$ in connection with a problem in algebraic number theory. The table in the present paper gives a comparison of the actual count of $P(N)$ with the value obtained from the right-hand side of (3), rounded off to the nearest integer.

TABLE

| $N$ | $P(N)$ | $0.761 \int_2^N (\log u)^{-2}\, du$ |
|---|---|---|
| 10 | 3 | 3 |
| 100 | 8 | 8 |
| 1000 | 23 | 26 |
| 5000 | 71 | 75 |
| 10000 | 117 | 123 |
| 20000 | 206 | 206 |
| 30000 | 275 | 281 |
| 40000 | 341 | 350 |
| 50000 | 404 | 416 |
| 60000 | 469 | 480 |
| 70000 | 532 | 542 |
| 80000 | 595 | 603 |
| 90000 | 649 | 662 |
| 100000 | 706 | 720 |
| 110000 | 758 | 777 |
| 113000 | 776 | 794 |

Needless to say, it would probably be worthwhile to attempt some numerical work in cases where at least one of the polynomials $f_1, f_2, \cdots, f_k$ has the property that a zero thereof does *not* generate an abelian extension of the field of rational numbers. An interesting example of this type would be the case of a single polynomial of the form $n^3 + a$, where $a$ is an integer which is not a perfect cube. Of course the determination of the numerical value of $C(f_1, f_2, \cdots, f_k)$ would be much more difficult in such a case.

It should be mentioned that the conjecture (1) and the result (2) may easily be extended to the case where $f_1, f_2, \cdots, f_k$ are merely integral-valued polynomials, that is, polynomials with rational coefficients which take integral values for all integral values of the variable. (For example, $\frac{1}{2}n^2 + \frac{1}{2}n + 1$ is such a polynomial.) The only change that must be made is that $\omega(p)/p$ in the formula for $C(f_1, f_2, \cdots, f_k)$ must be replaced by something a little more complicated, as we now explain. If $p$ is a prime and $m$ is a positive integer, let $\omega(p, m)$ denote the number of integers $j$ between 1 and $p^m$ inclusive such that $f_1(j)\, f_2(j) \cdots f_k(j)$ is a multiple of $p$. For given $p$, the ratio $\omega(p, m)/p^m$ has a value $\rho(p)$ independent of $m$ when $m$ is sufficiently large (specifically, when $p^m$ is larger than the highest power of $p$ dividing the denominator of any coefficient in any of the polynomials $f_i$). We need only replace $\omega(p)/p$ in the formula for $C(f_1, f_2, \cdots, f_k)$ by $\rho(p)$ in order to generalize (1) and (2) to arbitrary integral-valued polynomials. In the case of either (1) or (2), the generalization is an immediate consequence of the special case given originally.

Finally we remark that the conjectural formula (1) may be regarded as a quantitative form of the Hypothesis H of A. Schinzel [9], [10].

University of Illinois
Urbana, Illinois

1. ATLE SELBERG, "On an elementary method in the theory of primes," *Norske Vid. Selsk. Forh. Trondheim*, v. 19, no. 18, 1947, p. 64–67.
2. PAUL T. BATEMAN & ROSEMARIE M. STEMMLER, "Waring's problem in algebraic number fields and primes of the form $(p^r - 1)/(p^d - 1)$," *Illinois J. Math.*, v. 6, 1962, p. 142–156.
3. G. H. HARDY & J. E. LITTLEWOOD, "Some problems of partitio numerorum; III: On the expression of a number as a sum of primes," *Acta Math.*, v. 44, 1923, p. 1–70.
4. D. H. LEHMER, "Tables concerning the distribution of primes up to 37 millions," 1957, deposited in the UMT file and reviewed in *MTAC* v. 13, 1959, p. 56–57.
5. A. E. WESTERN, "Note on the number of primes of the form $n^2 + 1$," *Proc. Cambridge Philos. Soc.*, v. 21, 1922, p. 108–109.
6. DANIEL SHANKS, "On the conjecture of Hardy and Littlewood concerning the number of primes of the form $n^2 + a$," *Math. Comp.*, v. 14, 1960, p. 321–332.
7. DANIEL SHANKS, "A note on Gaussian twin primes," *Math. Comp.*, v. 14, 1960, p. 201–203.
8. DANIEL SHANKS, "On numbers of the form $n^4 + 1$," *Math. Comp.*, v. 15, 1961, p. 186–189.
9. A. SCHINZEL & W. SIERPIŃSKI, "Sur certaines hypothèses concernant les nombres premiers," *Acta Arith.*, v. 4, 1958, p. 185–208.
10. A. SCHINZEL, "Remarks on the paper 'Sur certaines hypothèses concernant les nombres premiers'," *Acta Arith.*, v. 7, 1961, p. 1–8.