

ON THE DISTRIBUTION OF PRIMES IN SHORT INTERVALS

P. X. GALLAGHER

One of the formulations of the prime number theorem is the statement that the number of primes in an interval $(n, n + h]$, averaged over $n \leq N$, tends to the limit λ , when N and h tend to infinity in such a way that $h \sim \lambda \log N$, with λ a positive constant.

In this note we study the distribution of values of $\pi(n + h) - \pi(n)$, for $n \leq N$ and $h \sim \lambda \log N$. We show that, assuming a certain uniform version of the (unproved) prime r -tuple conjecture of Hardy and Littlewood [3], the distribution tends to the Poisson distribution with parameter λ as $N \rightarrow \infty$. Using a sieve upper bound for the r -tuple problem, we also get an unconditional exponential upper bound for the tail of the distribution.

Our method has many features in common with the argument by which Hooley [4] has studied the distribution of values of the differences between consecutive integers prime to n , for $n/\phi(n)$ large. An analogous result for primes has been announced by Hooley in [5].

Explicitly, the r -tuple conjecture is an asymptotic formula for the number $\pi_{\mathbf{d}}(N)$ of positive integers $n \leq N$ for which $n + d_1, \dots, n + d_r$ are all prime. Here d_1, \dots, d_r are distinct integers. The formula is

$$\pi_{\mathbf{d}}(N) \sim \mathcal{S}_{\mathbf{d}} \frac{N}{\log^r N} \quad (N \rightarrow \infty), \quad (1)$$

provided $\mathcal{S}_{\mathbf{d}} \neq 0$, where

$$\mathcal{S}_{\mathbf{d}} = \prod_p \frac{p^{r-1}(p - v_{\mathbf{d}}(p))}{(p-1)^r},$$

and where $v_{\mathbf{d}}(p)$ is the number of distinct residue classes mod p occupied by d_1, \dots, d_r .

Formula (1) is the prime number theorem, for $r = 1$. For $r \geq 2$, it has not been proved for any \mathbf{d} ; the source of (1) in these cases is a heuristic application of the circle method, and a summation of the corresponding (multiple) singular series [3]. Lavrik [8] has proved that (1) holds in mean over cubes $1 \leq d_1, \dots, d_r \leq H$, in the range $N/\log^c N \leq H \leq N$; a similar mean result for the (small) cubes of side h would suffice for our purpose.

THEOREM 1. *Denote by $P_k(h, N)$ the number of integers $n \leq N$ for which the interval $(n, n + h]$ contains exactly k primes. Then*

$$P_k(h, N) \sim N \frac{e^{-\lambda} \lambda^k}{k!} \quad (2)$$

for $N \rightarrow \infty$, $h \sim \lambda \log N$, provided, for each r , (1) holds, uniformly for $1 \leq d_1, \dots, d_r \leq h$, with d_1, \dots, d_r distinct and $\mathcal{S}_{\mathbf{d}} \neq 0$.

Our argument for (2) goes through a computation of the moments of $\pi(n + h) - \pi(n)$, and depends on the fact that, for each r , \mathcal{S}_d averages to 1 over cubes:

$$\sum_{\substack{1 \leq d_1, \dots, d_r \leq h \\ \text{distinct}}} \mathcal{S}_d \sim h^r \quad (h \rightarrow \infty). \tag{3}$$

For $r = 2$, a smoothed variant of this was used by Hardy and Littlewood to refute earlier asymptotic Goldbach conjectures. A simple proof of (3) for $r = 2$, starting with the singular series representation for \mathcal{S}_d , was given by Bombieri and Davenport in [1]. Our proof of (3) starts with the product definition of \mathcal{S}_d , and is closer to an argument of Hooley in [5].

Using Selberg’s sieve, Klimov [7] obtained for each r the upper bound†

$$\pi_d(N) \lesssim 2^r r! \mathcal{S}_d \frac{N}{\log^r N} \tag{4}$$

for $N \rightarrow \infty$, uniformly for d in small cubes. For this, see Halberstam and Richert [2], Theorem 5.7. Using (4) instead of (1), we get upper bounds for the k th moments of $\pi(n + h) - \pi(n)$ for $n \leq N$, as Bombieri and Davenport did for $k = 2$. For large k , these bounds give

THEOREM 2. *For positive constants $\mu \geq \lambda \geq 1$, the number of $n \leq N$ for which $\pi(n + \lambda \log N) - \pi(n) > \mu$ is $\lesssim Ne^{-C\mu/\lambda}$, where C is an absolute constant.*

1. *Reduction to (3).* For each positive integer k ,

$$\begin{aligned} \sum_{n \leq N} (\pi(n + h) - \pi(n))^k &= \sum_{n \leq N} \sum_{\substack{n < p_1, \dots, p_k \leq n+h}} 1 \\ &= \sum_{r=1}^k \sigma(k, r) \sum \pi_{d_1, \dots, d_r}(N), \end{aligned}$$

where the inner sum is over all r -tuples d_1, \dots, d_r satisfying $1 \leq d_1 < \dots < d_r \leq h$, and $\sigma(k, r)$ is the number of maps from the set $\{1, \dots, k\}$ onto $\{1, \dots, r\}$. For the d with $\mathcal{S}_d \neq 0$, we use (1); for the others, d_1, \dots, d_r occupy all residue classes modulo some prime, so $\pi_d(N) \leq r$. Using (3), it follows that

$$\sum \pi_{d_1, \dots, d_r}(N) \sim \frac{h^r}{r!} \frac{N}{\log^r N},$$

and hence

$$\frac{1}{N} \sum_{n=1}^N (\pi(n + h) - \pi(n))^k \rightarrow m_k(\lambda), \tag{5}$$

with

$$m_k(\lambda) = \sum_{r=1}^k \sigma(k, r) \frac{\lambda^r}{r!}.$$

In §3, it is shown that $m_k(\lambda)$ is the k th moment of the Poisson distribution with

† The notation $F \lesssim G$ stands for $\overline{\lim} F/G \leq 1$.

parameter λ , and that the corresponding moment generating function is entire. The result (2) now follows from general theorems on moments [6, Chapter 4].

Putting $h = \lambda \log N$, and using (4) instead of (1), we get

$$\sum \pi_{d_1, \dots, d_r}(N) \lesssim (2\lambda)^r N,$$

from which it follows that

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N (\pi(n+h) - \pi(n))^k &\lesssim \sum_{r=1}^k \sigma(k, r)(2\lambda)^r \\ &\leq k(2\lambda k)^k. \end{aligned}$$

Hence the proportion of $n \leq N$ for which $\pi(n+h) - \pi(n) \geq \mu$ is $\lesssim k(2k\lambda/\mu)^k$. If $\mu/\lambda \geq 4$, we choose $k = \lceil \frac{1}{4}(\mu/\lambda) \rceil$. Then $k \geq \frac{1}{8}(\mu/\lambda)$, so the proportion is

$$\lesssim k2^{-k} \leq e^{-c\mu/\lambda}.$$

If $\mu/\lambda < 4$, the result is trivial.

2. Proof of (3). Let

$$D_{\mathbf{d}} = \prod_{i < j} (d_i - d_j).$$

Then $1 \leq v_{\mathbf{d}}(p) \leq r$, with equality at the right, unless $p \mid D_{\mathbf{d}}$. The p th factor in $\mathcal{S}_{\mathbf{d}}$ is

$$1 + \frac{p^r - v_{\mathbf{d}}(p)p^{r-1} - (p-1)^r}{(p-1)^r} = 1 + a(p, v_{\mathbf{d}}(p)), \tag{6}$$

where

$$a(p, v) \ll_r \begin{cases} (p-1)^{-2}, & v = r; \\ (p-1)^{-1}, & v < r. \end{cases} \tag{7}$$

It follows that the product for $\mathcal{S}_{\mathbf{d}}$ converges. Defining $a_{\mathbf{d}}(q)$ for squarefree q by

$$a_{\mathbf{d}}(q) = \prod_{p \mid q} a(p, v_{\mathbf{d}}(p)),$$

we get an absolutely convergent series expansion

$$\mathcal{S}_{\mathbf{d}} = \sum_q a_{\mathbf{d}}(q), \tag{8}$$

where the sum is over squarefree q .

We need an estimate for the remainder in (8) which is uniform for \mathbf{d} in the h -cube. It follows from the bounds on $a(p, v)$ that

$$\sum_{q > x} |a_{\mathbf{d}}(q)| \leq \sum_{q > x} \frac{\mu^2(q)C^{\omega(q)}}{\phi^2(q)} \phi((q, D)),$$

where $\omega(q)$ is the number of prime factors of q , and C is a positive constant depending only on r . Putting $q = de$ with $d \mid D$ and $(e, D) = 1$, this is

$$\begin{aligned} \sum_{d \mid D} \frac{\mu^2(d)C^{\omega(d)}}{\phi(d)} \sum_{\substack{e > x/d \\ (e, D) = 1}} \frac{\mu^2(e)C^{\omega(e)}}{\phi^2(e)} &\ll \sum_{d \mid D} \frac{\mu^2(d)C^{\omega(d)}}{\phi(d)} \frac{d}{x} \log^B x \\ &\ll (xh)^E/x, \end{aligned}$$

with a constant depending only on r and ε . It follows that

$$\sum_{\substack{d_1, \dots, d_r \leq h \\ \text{distinct}}} \mathcal{S}_a = \sum_{q \leq x} \sum_{\substack{d_1, \dots, d_r \leq h \\ \text{distinct}}} a_a(q) + O(h^r((xh)^\varepsilon/x)), \tag{9}$$

with a constant depending only on r and ε .

The inner sum in (9) is

$$\sum_v \prod_{p|q} a(p, v(p)) \{ \sum' 1 + O(h^{r-1}) \},$$

where $\sum' 1$ stands for the number of r -tuples of not necessarily distinct integers d_1, \dots, d_r with $1 \leq d_1, \dots, d_r \leq h$ which, for each prime $p|q$, occupy exactly $v(p)$ residue classes mod p ; the outer sum is over all "vectors" $= (\dots, v(p), \dots)_{p|q}$ with components satisfying $1 \leq v(p) \leq p$. A simple lattice point argument using the Chinese remainder theorem gives, for $q \leq h$,

$$\sum' 1 = \left\{ \left(\frac{h}{q}\right)^r + O\left(\frac{h}{q}\right)^{r-1} \right\} \prod_{p|q} \binom{p}{v(p)} \sigma(r, v(p));$$

the product representing the number of ways of choosing the residue classes of d_1, \dots, d_r mod q subject to the congruence restrictions in \sum' .

Thus the inner sum in (9) is

$$\left(\frac{h}{q}\right)^r A(q) + O\left(\left(\frac{h}{q}\right)^{r-1} B(q)\right) + O(h^{r-1} C(q)), \tag{10}$$

with

$$A(q) = \sum_v \prod_{p|q} a(p, v(p)) \binom{p}{v(p)} \sigma(r, v(p)),$$

$$B(q) = \sum_v \prod_{p|q} |a(p, v(p))| \binom{p}{v(p)} \sigma(r, v(p)),$$

$$C(q) = \sum_v \prod_{p|q} |a(p, v(p))|.$$

We have

$$A(q) = \prod_{p|q} \left\{ \sum_{v=1}^p a(p, v) \binom{p}{v} \sigma(r, v) \right\},$$

$$B(q) = \prod_{p|q} \left\{ \sum_{v=1}^p |a(p, v)| \binom{p}{v} \sigma(r, v) \right\},$$

$$C(q) = \prod_{p|q} \left\{ \sum_{v=1}^p |a(p, v)| \right\}.$$

We show first that $A(q) = 0$ for $q > 1$. Using (6), the p th factor in $A(q)$ is

$$(p-1)^{-r} \left\{ (p^r - (p-1)^r) \sum_{v=1}^p \binom{p}{v} \sigma(r, v) - p^{r-1} \sum_{v=1}^p v \binom{p}{v} \sigma(r, v) \right\}.$$

By formulae (i) and (ii) of §3, the two sums here are p^r and $p^{r+1} - (p - 1)^r p$ respectively, and the factor vanishes.

Using the bounds (7) for $a(p, v)$, we may estimate $B(q)$ and $C(q)$. By (i) of §3, the p th factor in $B(q)$ is $\ll p^r/(p - 1)$, so

$$B(q) \ll C^{\omega(q)} \frac{q^r}{\phi(q)}.$$

More simply, the p th factor in $C(q)$ is $\ll p/(p - 1)$, so

$$C(q) \ll C^{\omega(q)} \frac{q}{\phi(q)}.$$

Returning to (9) and (10), it follows that (9) is h^r plus a remainder term which is

$$\begin{aligned} &\ll h^{r-1} \sum_{q \leq x} C^{\omega(q)} \frac{q}{\phi(q)} + h^r (xh)^\epsilon/x \\ &\ll h^{r-1} x^{1+\epsilon} + h^r (hx)^\epsilon/x \\ &\ll h^{r-\frac{1}{2}+\epsilon}, \end{aligned}$$

choosing $x = h^{\frac{1}{2}}$. Since $x \leq h$, the conditions $q \leq h$, assumed earlier, are satisfied.

3. *Combinatorial identities.* We prove here the standard identities for the “Stirling numbers of the second kind” $\sigma(k, r)/r!$ which have been used above. These are

- (i) $\sum_{v=1}^p \binom{p}{v} \sigma(r, v) = p^r,$
- (ii) $\sum_{v=1}^p v \binom{p}{v} \sigma(r, v) = p^{r+1} - (p - 1)^r p,$
- (iii) $\sum_{v=1}^r \sigma(r, v) \frac{\lambda^v}{v!} = \sum_{p=0}^{\infty} p^r \frac{e^{-\lambda} \lambda^p}{p!},$
- (iv) $\sum_{r=0}^{\infty} \frac{m_r(\lambda) z^r}{r!} = e^{-\lambda} e^{\lambda e^z};$

the last two identities show that $m_r(\lambda)$, the left side of (iii), is the r th moment of the Poisson distribution with parameter λ , and that the corresponding moment generating function (iv) is entire.

To prove (i), classify the maps from $\{1, \dots, r\}$ to $\{1, \dots, p\}$ by the size of the image. There are $\binom{p}{v}$ subsets of size v in $\{1, \dots, p\}$; for each such subset, the number of maps with this image is $\sigma(r, v)$. To prove (ii), write

$$\binom{p}{v} = p \binom{p-1}{v-1} = p \binom{p}{v} - p \binom{p-1}{v},$$

and use (i). To prove (iii), multiply (i) by $\lambda^p/p!$ and sum over p :

$$\sum_{v=1}^r \sigma(r, v) \sum_{p=0}^{\infty} \binom{p}{v} \frac{\lambda^p}{p!} = \sum_{p=0}^{\infty} p^r \frac{\lambda^p}{p!}.$$

From this and

$$\sum_{p=0}^{\infty} \binom{p}{v} \frac{\lambda^p}{p!} = \frac{\lambda^v}{v!} e^{\lambda},$$

the identity (iii) follows. To prove (iv), multiply (iii) by $z^r/r!$ and sum over r .

References

1. E. Bombieri and H. Davenport. "Small differences between prime numbers", *Proc. Roy. Soc. Ser. A.* 293 (1966), 1–18.
2. H. Halberstam and E. Richert. *Sieve Methods* (Academic Press, 1974).
3. G. H. Hardy and J. E. Littlewood. "Some problems of 'Partitio Numerorum': III. On the expression of a number as a sum of primes", *Acta Mathematica*, 44 (1922), 1–70.
4. C. Hooley. "On the difference between consecutive numbers prime to n . II", *Publ. Math. Debrecen*, 12 (1965), 39–49.
5. C. Hooley. "On the intervals between consecutive terms of sequences", *Proc. Symp. Pure Math.*, 24 (1973), 129–140.
6. M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics, Vol. I* (Griffin, 1958).
7. N. I. Klimov. "Combination of elementary and analytic methods in the theory of numbers", *Uspehi Mat. Nauk*, 13 (1958), no. 3 (81), 145–164.
8. A. F. Lavrik. "On the theory of distribution of primes, based on I. M. Vinogradov's method of trigonometrical sums", *Trudy Mat. Inst. Steklov.*, 64 (1961), 90–125.

Mathematics Department
Columbia University
New York, New York 10027 (U.S.A.).

10H15: NUMBER THEORY; *Multiplicative Theory; Distribution of primes.*

Received on the 10th of October, 1975.