ACTA ARITHMETICA
125.2 (2006)

Primes with preassigned digits

by

Glyn Harman (Egham)

1. Introduction. In 1951 Sierpiński [11] began the investigation of prime numbers with preassigned digits. In [12] he showed that in any given base g and with given a, b with $1 \le a \le g-1$, $\gcd(b,g)=1$, $1 \le b \le g-1$, one can find infinitely many primes p having a as its first digit and b as its last. This result was extended to arithmetic progressions of primes in [4]. As we show in the next section, these results are elementary deductions from deeper results on the distribution of primes. Indeed, one can make certain quantitative statements about how many initial or final digits one can prescribe. Recently Wolke [13] has considered the more difficult case where one preassigns at most two digits *anywhere* in the expansion of an integer with k digits, and obtains an asymptotic formula (valid for $k \to \infty$) in this case.

We need some notation to state the problem. Given positive integers g, t, k we shall say that a sequence in $\mathbb{Z}^2$ is (g,t,k)-*admissible* if it is of the form $\mathbf{d}_1, \dots, \mathbf{d}_t$ with $\mathbf{d}_j = (n_j, a_j)$ where $1 \le n_1 < \cdots < n_t \le k$ and $a_j \in \{0, \dots, g-1\}$ with $\gcd(a_t, g) = 1$ if $n_t = k$, and $a_1 \ge 1$ if $n_1 = 1$. Wolke makes a quantitative conjecture, which in a qualitative form reads as follows.

Conjecture. *Let $g \ge 2$, $t \ge 1$ be given. Then there exists a positive quantity $K = K(g,t)$ such that for all $k \ge K$, and any (g,t,k)-admissible sequence, there are primes $p = b_1 \dots b_k$ in base g with $b_{n_j} = a_j$, $1 \le j \le t$.*

Wolke gives a proof of this result dependent on the Generalised Riemann Hypothesis. It is the purpose of this paper to establish this conjecture unconditionally.

2. Sketch of proof. First we distinguish the easy and difficult parts of the problem. By the main theorem in [2], for any large x there are primes in the interval $[x, x + x^{0.525}]$. It follows that for large enough k we can preassign the first $0.475k - 1$ digits. Even with the Riemann Hypothesis we

2000 *Mathematics Subject Classification*: Primary 11N05.

could not expect to do better than preassigning the first $(1/2-\varepsilon)k$ digits (although here we could take $\varepsilon \to 0$ as $k \to \infty$). The strongest possible conjecture on gaps between primes would increase this to $(1-\varepsilon)k$ digits. Now the question of preassigning the final digits corresponds to the difficult problem of primes in arithmetic progressions. One would therefore expect the maximum number of final digits that can be preassigned to depend on the best currently known value for Linnik's constant. We note that Heath-Brown [9] has established that for all large q, given a coprime to q there is a prime $p \equiv a \pmod q$ with $p \le q^{11/2}$. This enables one to preassign the last $2k/11$ digits essentially.

One can make some attempt to try to combine these results to preassign blocks of digits at the begining and end of the expansion of an integer, but progress is difficult without good zero-free regions for L-functions. In the case of working to base $g = p^c$ for some c, with p an odd prime (and hence the arithmetic progressions considered are to modulus p^r for some r), the work of Gallagher [6] can be brought into play. Using the techniques outlined in [8] it is then possible to preassign $0.472k$ digits, so long as they appear in two continuous blocks: one at the start and one at the end of the expansion. The challenging problem is to consider the preassigned digits spread throughout the expansion and here we present a method which does this.

To simplify notation we shall henceforth work only in base 10, but the proof works identically for all bases. We write $\{\ \}$ to denote the fractional part of a real number. We note that if, in base 10, we have $p = b_1 \dots b_k$, then $b_r = a$ is equivalent to

$$\{p/10^u\} \in [a/10, (a+1)/10), \tag{1}$$

where $u = k+1-r$. There are well known techniques for picking out such a fractional part condition using exponential sums; see [1, Chapter 2] for example. The consideration of (1) would require the right sort of bounds on

$$\sum_{l=1}^{L} \Big| \sum_{p<10^k} e(lp10^{-u}) \Big|. \tag{2}$$

Indeed, by [1, Theorem 2.2], we obtain a solution if, with $L = 20$, the above sum is $< \frac{1}{6}\pi(10^k)$. Here $\pi(x)$ denotes the number of primes up to x (or if one wanted a "genuine" k digit prime then replace $\pi(10^k)$ with $\pi(10^k) - \pi(10^{k-1})$).

Now the estimation of such exponential sums is well known in analytic number theory (see, for example, [5, Chapter 25]). For a successful outcome one essentially requires 10^u to be neither near 10^k in size, nor near 1. To be precise, one requires a condition (for large k) like $10^k k^{-4} > 10^u > k^4$. Of

course, if 10^u were to be outside this range then we could handle the situation using either primes in short intervals or primes in arithmetic progressions.

So, if we had an estimate for exponential sums over primes in short intervals in arithmetic progression then we would expect to be able to settle the conjecture for $t = 3$. The argument would go as follows. If all three digits were near the beginning then use primes in short intervals. If they were all near the end then use primes in arithmetic progressions. If there were a 1–2 or 2–1 split between these cases then it can be handled by modifying the smallest prime in an arithmetic progression result (a weaker exponent would have to be used here—indeed the Siegel–Walfisz theorem suffices). In the other cases we need to bring a bound for an expression like (2) into play (with restrictions on p). This is provided by the following result.

LEMMA 1. *Suppose that* $1 < y < x$, $\gcd(a, q) = \gcd(b, d) = 1$, $\gcd(d, q) = h$. *Then we have*

$$\sum_{\substack{y\le p<x\\ p\equiv b\,(\mathrm{mod}\,d)}} e\Big(\frac{ap}{q}\Big) \ll x(\log x)^2\Big(\frac{h}{dq^{1/2}} + \Big(\frac{q}{xh}\Big)^{1/2} + \frac{1}{x^{1/5}d^{2/5}}\Big). \tag{3}$$

Proof. This follows (after partial summation) by applying the theorem in [3] with $N = x$ and $N = y$. If y is very near x then one could give a sharper bound, but that will not be required here. ∎

In applying the above result to the case $t = 3$ with a digit near the start and end of the expansion fixed we will have $a = l(l, 10^u)^{-1}$, $q = 10^u(l, 10^u)^{-1}$, d will be a small power of 10 and h will be d. The challenge still remains to deal with more than one digit at neither end of the expansion. To do this we need to treat more than one condition like (1) simultaneously. Again, there are ways to do this using exponential sums. Before the problem can be transformed, however, there is a possible natural difficulty to overcome. If two or more digits are adjacent then it would be most sensible to treat them as a block. For example

$$\{p10^{-r}\} \in [a_1/10, (a_1 + 1)/10) \quad \text{and} \quad \{p10^{-(r+1)}\} \in [a_2/10, (a_1 + 1)/10)$$

would become

$$\{p10^{-(r+1)}\} \in [(a_1/10 + a_2)/10, (a_1/10 + a_2 + 1)/10).$$

The interference between consecutive or very near digits is translated into the exponential sum estimates as the expression

$$\frac{l_1}{10^r} + \frac{l_2}{10^{r+1}}$$

becoming zero or a fraction with very small denominator for relatively low values of l_1, l_2, namely $10l_1 = -l_2$.

The reader should by now see that it ought to be possible to tackle the case $t = 3$. The case for larger t just becomes more complicated. We need to determine when we are treating near digits as a "block", and when to consider them separately. Also those digits close to either end need to be isolated from those considered via the above fractional part argument. The principle we began to expound in the case $t = 3$ remains, however, as shall be seen in the next section.

3. Completion of the proof. First we state the lemma converting the problem from simultaneous Diophantine approximation to exponential sum form.

LEMMA 2. *Let $\mathcal{I}_j$ be subintervals of $[0,1)$, $1 \le j \le u$, with $|\mathcal{I}_j| = 2uL_j^{-1}$. Let $r_m \in \mathbb{R}$ for $1 \le m \le X$ and $\alpha_j \in \mathbb{R}$ for $1 \le j \le u$. Suppose that*

$$\sum_{\substack{|l_j| \le L_j \\ \max |l_j| > 0}} \Big| \sum_{m \le X} e(r_m(\alpha_1 l_1 + \cdots + \alpha_u l_u)) \Big| \le \frac{X}{4u^2 - 1}. \tag{4}$$

Then there is a solution to

$$\{r_m \alpha_j\} \in \mathcal{I}_j, \quad 1 \le j \le u.$$

Proof. This follows from [7, Lemma 5]. ∎

We also need a form of the Siegel–Walfisz theorem to count the number of primes in short intervals restricted to an arithmetic progression.

LEMMA 3. *Let $N \ge 1$ be given. Let x, q, y be related by*

$$1 \le q \le (\log x)^N, \quad (x - y)(\log x)^N \ge x.$$

Suppose $\gcd(a, q) = 1$. *Then the number of primes $p \equiv a \pmod q$ with $y \le p \le x$ is*

$$> \frac{x - y}{2\phi(q) \log x}, \tag{5}$$

for all $x > x_0(N)$.

Proof. This follows from [5, p. 133], for example. The value $x_0(N)$ is ineffective owing to the appeal to Siegel's theorem. ∎

Proof of Conjecture. We have not tried to optimise the argument; there are various ways it could be made more efficient with more effort. We will not fix the relation between k and t initially, although the reader should bear in mind that k will eventually be chosen to be much larger than t. We can therefore start by assuming that $k > 8^{4t}$, which suffices for Lemma 4, for example. In the following, $\log_r z$ denotes the logarithm of z to base r. Let $A > 1$ be a fixed constant to be determined later, and write $\kappa = \log_8(A \log_{10} k)$. First

we isolate the preassigned digits amongst the initial and final digits (which will be considered using primes in short intervals in arithmetic progressions) from those which will be tackled *via* the Diophantine approximation argument.

LEMMA 4. *With the above notation, there must be a j with $\kappa \le j \le t+1+\kappa$ such that there are no prescribed digits in place n with either*

$$8^j < n \le 8^{j+1} \quad \text{or} \quad k - 8^{j+1} < n \le k - 8^j. \tag{6}$$

Proof. There are at least $t+1$ disjoint pairs of intervals for n defined by (6) with j in the stated interval, and only t numbers we wish to avoid. ∎

When applying the above lemma, in case of ambiguity we take the smallest such j. We now fix the first and last 8^j digits, say by choosing to be 1 all those digits not already preassigned. We note that

$$k^A \le 10^{8^j} \le k^{8^{t+2}A}.$$

Write $V = 10^{8^j}$. It follows that V is of size $(\log 10^k)^B$ for some B bounded in terms of t. Putting this another way, $k^\alpha < V < k^\beta$, where α, β depend at most on t. This will be crucial since we need $k^N > V$ for some fixed N in order to apply Lemma 3, and we shall eventually choose $A = 3$ so that $k^3 \le V$. By Lemma 3 the number of primes with the first and last 8^j digits fixed, say X, is

$$\gg \frac{10^{k-2\cdot 8^j}}{k} = \frac{10^k}{kV^2}. \tag{7}$$

We write $\mathcal{A}$ for this set of primes, and also note for future reference that $10^{8^{j+1}} = V^8$.

We now divide the remaining digits into blocks according to where the preassigned digits fall. Let the remaining n_j be $m_1 < \cdots < m_v$. We form blocks $\mathcal{B}_1, \ldots, \mathcal{B}_u$ using the rule

$$m_j \in \mathcal{B}_r \text{ and } m_{j+1} < m_j + t \;\Rightarrow\; m_{j+1} \in \mathcal{B}_r.$$

We then have at most t blocks and each block encompasses at most t^2 digits. We then fix all previously non-assigned digits in each block (say by putting them all equal to zero). We thus have at most t simultaneous conditions that p must satisfy of the form

$$\{p/10^{k-f_j+1}\} \in \mathcal{I}_j,$$

where f_j is the first position of a block (so it equals m_h for some h) and $\mathcal{I}_j$ has length 10^{-s_j} where s_j is the length (that is, the number of digits) of the block. Say we have u blocks altogether. By Lemma 2 we must then show

that

$$\sum_{\substack{|l_j|\le L_j\\ \max|l_j|>0}} \Big|\sum_{p\in\mathcal{A}} e(p(\alpha_1 l_1+\cdots+\alpha_u l_u))\Big| \le \frac{X}{4u^2-1}, \tag{8}$$

where $L_j = 2u10^{s_j}$, and $\alpha_j = 10^{-f_j}$.

Now let

$$\alpha_1 l_1+\cdots+\alpha_u l_u = a/q \quad \text{with } \gcd(a,q)=1.$$

By our construction of the blocks,

$$2L_j\alpha_j = 4u10^{s_j-f_j} \le 4u10^{-t-f_{j+1}} < \alpha_{j+1}.$$

It follows that $a\neq 0$, since if w is the largest value with $l_w\neq 0$, then $|l_w\alpha_w|$ exceeds the sum of all the moduli of all the other terms. Also

$$2L_u\alpha_u \le 4u10^{-8^{j+1}} \le 4tV^{-8}.$$

Thus

$$q > V^{-7},$$

assuming $k^A > 4t$ (we have already assumed k to be much larger than this in fact!). We also have

$$q < 10^{k-8^{j+1}} = 10^k V^{-8}.$$

Now the number of values taken by $(l_1,\dots,l_u)$ in (8) is

$$< \prod_{j=1}^{u}(2L_j+1) < \prod_{j=1}^{u}(3u10^{s_j}) \le (3t)^t 10^{t^2}.$$

We can therefore apply Lemma 1 (we note that $d < q^{1/8}$ by our construction, but we only need the bounds $1\le h\le d$) to give a bound for the left-hand side of (8), which is

$$\begin{aligned} &< (3t)^t 10^{t^2} 10^k k^2\left(\frac{1}{q^{1/2}}+\left(\frac{q}{10^k}\right)^{1/2}+\frac{1}{10^{k/5}}\right)\\ &\ll (3t)^t 10^{t^2} 10^k k^2 V^{-7/2} \ll X\left(\frac{(3t)^t 10^{t^2} k^3}{V^{3/2}}\right), \end{aligned}$$

using (7). Hence (8) is satisfied (with $A=3$ say) for all large k and the proof is complete. To be more precise, we need to choose k so that

$$(3t)^t 10^{t^2} < k,$$

and that it is "sufficiently large" for Lemma 3 to operate—and this makes the dependence on t ineffective.

We leave it as an exercise for the reader to show that one can obtain an asymptotic formula for the number of primes as $k\to\infty$. In this case one might want to commence with Vinogradov's familiar construction of a

trigonometric polynomial approximating the characteristic function of an interval (mod 1), as in [1, Chapter 2.1]. Also, one needs to sum over all possible combinations of digits that were fixed in the argument but were not among those preassigned.

References

[1] R. C. Baker, *Diophantine Inequalities*, London Math. Soc. Monogr. (N.S.) 1, Oxford Univ. Press, New York, 1986.

[2] R. C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes, II*, Proc. London Math. Soc. (3) 83 (2001), 532–562.

[3] A. Balog and A. Perelli, *Exponential sums over primes in an arithmetic progression*, Proc. Amer. Math. Soc. 93 (1985), 578–582.

[4] L. J. Borucki and J. B. Díaz, *A note on primes, with arbitrary initial or terminal decimal ciphers, in Dirichlet arithmetic progressions*, Amer. Math. Monthly 81 (1974), 1001–1002.

[5] H. Davenport, *Multiplicative Number Theory*, 2nd ed., Springer, New York, 1980.

[6] P. X. Gallagher, *Primes in progressions to prime-power modulus*, Invent. Math. 16 (1972), 191–201.

[7] G. Harman, *Small fractional parts of additive forms*, Philos. Trans. Roy. Soc. London Ser. A 345 (1993), 327–338.

[8] —, *On the number of Carmichael numbers up to x*, Bull. London Math. Soc. 37 (2005), 641–650.

[9] D. R. Heath-Brown, *Zero-free regions for Dirichlet L-functions, and the least prime in an arithmetic progression*, Proc. London Math. Soc. (3) 64 (1992), 265–338.

[10] A. F. Lavrik, *An analytic method of estimating trigonometric sums over primes in an arithmetic progression*, Dokl. Akad. Nauk SSSR 248 (1979), 1059–1063 (in Russian); English transl.: Soviet Math. Dokl. 20 (1979), 1121–1124.

[11] W. Sierpiński, *Sur l'existence des nombres premiers avec une suite arbitraire de chiffres initiaux*, Matematiche (Catania) 6 (1951), 135–137.

[12] —, *Sur les nombres premiers ayant des chiffres initiaux et finals donnés*, Acta Arith. 5 (1959), 265–266.

[13] D. Wolke, *Primes with preassigned digits*, ibid. 119 (2005), 201–209.

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
E-mail: G.Harman@rhul.ac.uk

Received on 21.10.2005
and in revised form on 12.6.2006 (5087)