Functiones et Approximatio
45.2 (2011), 231–253

CONSTRUCTION OF NORMAL NUMBERS BY CLASSIFIED PRIME DIVISORS OF INTEGERS

Jean-Marie De Koninck, Imre Kátai

Abstract: Given an integer $d \geqslant 2$, a *d-normal number*, or simply a *normal number*, is a real number whose d-ary expansion is such that any preassigned sequence, of length $k \geqslant 1$, of base d digits from this expansion, occurs at the expected frequency, namely $1/d^k$. We construct large families of normal numbers using classified prime divisors of integers.

Keywords: normal numbers, primes, shifted primes, arithmetic function

1. Introduction

Given an integer $d \geqslant 2$, a *d-normal number*, or simply a *normal number*, is a real number whose d-ary expansion is such that any preassigned sequence, of length $k \geqslant 1$, of base d digits from this expansion, occurs at the expected frequency, namely $1/d^k$. Equivalently, given a positive real number $\eta < 1$ whose expansion is $\eta = 0, a_1 a_2 \ldots$, where $a_i \in \{0, 1, \ldots, d-1\}$, that is, $\eta = \sum_{j=1}^{\infty} \frac{a_j}{d^j}$, we say that η is a normal number if the sequence $\{d^m \eta\}$, $m = 1, 2, \ldots$ (here $\{y\}$ stands for the fractional part of y), is uniformly distributed in the interval $[0, 1[$. Clearly, both definitions are equivalent, because the fact that the sequence $\{d^m \eta\}$, $m = 1, 2, \ldots$, is uniformly distributed in $[0, 1[$ occurs if and only if for every integer $k \geqslant 1$ and $b_1 \ldots b_k \in \{0, 1, \ldots, d-1\}^k$, we have

$$\lim_{N \to \infty} \frac{1}{N} \#\{j < N : a_{j+1} \ldots a_{j+k} = b_1 \ldots b_k\} = \frac{1}{d^k}.$$

Identifying if a given real number is a normal number is not an easy task. For instance, classical numbers such as π, e and $\sqrt{2}$ have not yet been proven to be normal numbers. Even constructing specific normal numbers is a no smaller challenge.

Several authors studied the problem of constructing normal numbers. One of the first was Champernowne [1] who, in 1933, was able to prove that the number

2010 Mathematics Subject Classification: primary: 11K16, 11N37, 11A41

made up of the concatenation of the natural numbers, namely the number

$$0,12345678910111213141516171819202l\ldots,$$

is normal in base 10. In 1946, Copeland and Erdős [2] showed that the same is true if one replaces the sequence of natural numbers by the sequences of primes, namely for the number

$$0,23571113171923293137\ldots$$

In the same paper, they conjectured that if $f(x)$ is any polynomial whose values at $x = 1, 2, 3, \ldots$ are positive integers, then the decimal $0, f(1)f(2)f(3)\ldots$, where $f(n)$ is written in base 10, is a normal number. In 1952, Davenport and Erdős [3] proved this conjecture. In 1997, Nakai and Shiokawa [9] showed that if $f(x)$ is any nonconstant polynomial taking only positive integral values for positive integral arguments, then the number $0, f(2)f(3)f(5)f(7)\ldots f(p)\ldots$, where p runs through the prime numbers, is normal. In 2008, Madritsch, Thuswaldner and Tichy [8] extended the results of Nakai and Shiokawa by showing that, if f is an entire function of logarithmic order, then the numbers

$$0,[f(1)]_q[f(2)]_q[f(3)]_q\ldots \qquad \text{and} \qquad 0,[f(2)]_q[f(3)]_q[f(5)]_q[f(7)]_q\ldots,$$

where $[f(n)]_q$ stands for the base q expansion of the integer part of $f(n)$, are normal.

In this paper, we explore another approach, by constructing large families of normal numbers using classified prime divisors of integers.

2. Notations

Given an integer $n \geqslant 2$, we let $\omega(n)$ stand for the number of distinct prime divisors of n and set $\omega(1) = 0$. We shall also write $p(n)$ and $P(n)$ for the smallest and largest prime factor of $n \geqslant 2$, respectively. As usual, φ stands for the Euler Function.

Throughout this paper, p and q, with or without subscripts, will always denote prime numbers. Given a particular set of primes $\mathcal{Q}$, we let $\mathcal{N}(\mathcal{Q})$ stand for the semi-group generated by the primes belonging to $\mathcal{Q}$. Moreover, at times, we shall write x_1 for $\log x$, and further define $x_{k+1} = \log x_k$ for $k = 1, 2, \ldots$

Given a real number $x \geqslant 2$ and coprime integers k, ℓ, we let $\pi(x; k, \ell)$ stand for the number of prime numbers $p \leqslant x$ such that $p \equiv \ell \pmod{k}$. For each real number $x \geqslant 2$, we set $\mathrm{li}(x) := \displaystyle\int_2^x \frac{dt}{\log t}$, a function often called the logarithmic integral. We will also be using the well known function

$$\Psi(x, y) := \#\{n \leqslant x : P(n) \leqslant y\} \qquad (2 \leqslant y \leqslant x).$$

Let $\wp$ stand for the set of all primes. Given an integer $d \geqslant 2$, we shall be interested in disjoint sets of prime numbers $\wp_0, \wp_1, \ldots, \wp_{d-1}$ such that

$$\wp = \mathcal{R} \cup \wp_0 \cup \wp_1 \cup \cdots \cup \wp_{d-1}, \tag{2.1}$$

where $\mathcal{R}$ is a given finite (perhaps empty) set of primes. Relation (2.1) is called a *disjoint classification of primes.* For instance, the sets $\wp_0 = \{p : p \equiv 1 \pmod 4\}$, $\wp_1 = \{p : p \equiv 3 \pmod 4\}$ and $\mathcal{R} = \{2\}$ provide a disjoint classification of primes.

For each positive integer d, let $A_d := \{0, 1, \dots, d-1\}$. Given an interval of real numbers I and a set of primes S, we write $\pi(I)$ for the number of prime numbers located in the interval I, while we write $\pi(I|S)$ for the number of primes $p \in S$ which belong to I.

Given an integer $t \geqslant 1$, an expression of the form $i_1 i_2 \dots i_t$, where each i_j is one of the numbers $0, 1, \dots, d-1$, is called a *word* of length t. We sometimes write $\lambda(\alpha) = t$ to indicate that α is a *word* of length t. At this point we introduce the symbol Λ to denote the *empty word* and the function $H : \wp \to A_d$ defined by

$$H(p) = \begin{cases} j & \text{if } p \in \wp_j \ (j = 0, 1, \dots, d-1), \\ \Lambda & \text{if } p \in \mathcal{R}, \end{cases}$$

and further extend the domain of the function H to all prime powers p^a by simply setting $H(p^a) = H(p)$.

Letting A_d^* be the set of finite words over A_d, we introduce the function $R : \mathbb{N} \to A_d^*$ defined as follows. If $n = p_1^{a_1} \cdots p_r^{a_r}$, where $p_1 < \dots < p_r$ are primes and each $a_i \in \mathbb{N}$, we set

$$R(n) = H(p_1) \dots H(p_r), \tag{2.2}$$

where on the right hand side of (2.2), we omit $H(p_i) = \Lambda$ if $p_i \in \mathcal{R}$. For convenience, we set $R(1) = \Lambda$.

In the example already mentioned above, that is, choosing $\wp_0 = \{p : p \equiv 1 \pmod 4\}$, $\wp_1 = \{p : p \equiv 3 \pmod 4\}$ and $\mathcal{R} = \{2\}$, we easily get that

$$\{R(1), R(2), \dots, R(15)\} = \{\Lambda, \Lambda, 1, \Lambda, 0, 1, 1, \Lambda, 1, 0, 1, 1, 0, 1, 10\}.$$

Now, consider the situation where $\wp = \mathcal{R} \cup \wp_o \cup \dots \cup \wp_{d-1}$ is a disjoint classification of primes, and let R be defined as in (2.2). Consider the number

$$\xi = 0, R(1)R(2)\dots,$$

which represents an infinite sequence over A_d and which in turn, by concatenating the finite words $R(1)$, $R(2)$, $\dots$, can be considered as the d-ary expansion of a real number, namely the real number ξ. In what follows, we shall examine what further conditions should be required in order to claim that the above number ξ is indeed a d-normal number.

3. Main results

Theorem 3.1. *Let $d \geqslant 2$ be an integer and let $\wp = \mathcal{R} \cup \wp_o \cup \dots \cup \wp_{d-1}$ be a disjoint classification of primes. Assume that, for a certain constant $c \geqslant 5$,*

$$\pi([u, u+v]|\wp_j) = \frac{1}{d}\pi([u, u+v]) + O\left(\frac{u}{\log^c u}\right)$$

uniformly for $2 \leqslant v \leqslant u$, $j = 0, 1, \ldots, d-1$, as $u \to \infty$. Further, let R be defined on $\mathbb{N}$ as in (2.2) and consider the number

$$\xi = 0, R(1)R(2)\ldots \tag{3.1}$$

Consider the right hand side of (3.1) as the d-ary expansion of a real number. Then ξ is a d-normal number.

Theorem 3.2. *Given two positive integers a and D such that $\gcd(D,a) = 1$, let $\wp_h := \{p : p \equiv h \pmod{D}\}$ for $\gcd(h,D) = 1$. Let $h_0, h_1, \ldots, h_{\varphi(D)-1}$ be those positive integers $< D$ which are relatively prime with D. Further let $\mathcal{R} = \{p : p|D\}$ and set*

$$R(p^a) = R(p) = \begin{cases} j & \text{if } p \equiv h_j \pmod{D} \\ \Lambda & \text{if } p|D. \end{cases}$$

Let ξ be the real number whose $\varphi(D)$-ary expansion is given by

$$\xi = 0, R(2+a)R(3+a)R(5+a)\ldots R(p+a)\ldots,$$

where $p + a$ is the sequence of shifted primes. Then ξ is a $\varphi(D)$-normal number.

Given a positive real number Y, then for each integer $n \geqslant 2$, let

$$A(n|Y) = \prod_{\substack{p^\alpha \| n \\ p \leqslant Y}} p^\alpha.$$

Theorem 3.3. *Let $a \neq 0$ be an integer. Let ε_x be a function which tends to 0 very slowly as $x \to \infty$, but such that $1/\varepsilon_x = o(\log\log x)$. Let $\mathcal{K}_x := \{K \in \mathbb{N} : P(K) \leqslant x^{\varepsilon_x}\}$. For each $K \in \mathcal{K}_x$, define*

$$\Delta_K(x) := \#\{p \leqslant x : A(p+a|x^{\varepsilon_x}) = K\}$$

and, for $\gcd(a,K) = 1$,

$$\begin{aligned} \kappa(K) &:= \prod_{\substack{p < x^{\varepsilon_x} \\ \gcd(p,Ka)=1}} \left(1 - \frac{1}{p-1}\right) \cdot \prod_{\substack{p|K \\ \gcd(p,a)=1}} \left(1 - \frac{1}{p}\right) \\ &= \prod_{\substack{p < x^{\varepsilon_x} \\ \gcd(p,Ka)=1}} \left(1 - \frac{1}{p-1}\right) \cdot \frac{\varphi(K)}{K}. \end{aligned} \tag{3.2}$$

Let also δ_x be a function satisfying $\lim_{x\to\infty} \delta_x = 0$ and $\lim_{x\to\infty} \delta_x/\varepsilon_x = +\infty$. Then,

$$\begin{aligned} \sum_{\substack{K \in \mathcal{K}_x,\ K < x^{\delta_x} \\ \gcd(K,a)=1}} \left|\Delta_K(x) - \frac{\kappa(K)}{\varphi(K)} li(x)\right| &\ll \exp\left\{-\frac{1}{2}\frac{\delta_x}{\varepsilon_x} \log \frac{\delta_x}{\varepsilon_x}\right\} \cdot \pi(x) \\ &\quad + O\left(\frac{x}{\log^A x}\right) + O(\varepsilon_x \pi(x)), \end{aligned} \tag{3.3}$$

where A is an arbitrary constant.

Moreover,

$$\lim_{x\to\infty}\frac{1}{\pi(x)}\sum_{\substack{K\in\mathcal{K}_x\\ \gcd(K,a)=1}}\left|\Delta_K(x)-\frac{\kappa(K)}{\varphi(K)}li(x)\right|=0. \tag{3.4}$$

Let $k\geqslant 1$ be a fixed integer and set $E(n):=n(n+1)\cdots(n+k-1)$. Moreover, for each positive integer n, define

$$e(n)=\prod_{\substack{q^\beta\,\|\,E(n)\\ q\leqslant k-1}} q^\beta.$$

We shall now define the sequence h_n on the prime powers q^a of $E(n)$ as follows:

$$h_n(q^a)=h_n(q)=\begin{cases}\Lambda & \text{if } q|e(n)\\ \ell & \text{if } q|n+\ell,\ \gcd(q,e(n))=1.\end{cases}$$

If $E(n)=q_1^{a_1}q_2^{a_2}\cdots q_r^{a_r}$ where $q_1<q_2<\cdots<q_r$ are primes an each $a_i\in\mathbb{N}$, then we set

$$S(E(n))=h_n(q_1)h_n(q_2)\cdots h_n(q_r).$$

Theorem 3.4. *Let k, E and S be as above. Let ξ be the real number whose k-ary expansion is given by*

$$\xi=0,S(E(1))S(E(2))\ldots S(E(n))\ldots \tag{3.5}$$

Then, ξ is a k-normal number.

Theorem 3.5. *Let $p_1<p_2<\cdots$ be the sequence of primes, and let k, E and S be as above. Let ξ be the real number whose k-ary expansion is given by*

$$\xi=0,S(E(p_1+1))S(E(p_2+1))\ldots$$

Then, ξ is a k-normal number.

4. Preliminary lemmas

Let w_x be a nondecreasing function which tends to $+\infty$ as $x\to\infty$. Let x be a large number.

Lemma 4.1. *Let $\alpha=i_1\ldots i_k\in A_d^k$ be an arbitrary word and let R be as in (2.2), and define*

$$N_k(Y|w_x):=\#\{p_1^{a_1}\cdots p_k^{a_k}\leqslant Y: w_x<p_1<\cdots<p_k\},$$
$$N_k(Y|w_x;\alpha):=\#\{p_1^{a_1}\cdots p_k^{a_k}\leqslant Y: w_x<p_1<\cdots<p_k,\ R(p_1^{a_1}\cdots p_k^{a_k})=\alpha\}.$$

Assume that, uniformly $2\leqslant v\leqslant u$, $j=0,\ldots,d-1$,

$$\pi([u,u+v]|\wp_j)=\frac{1}{d}\pi([u,u+v])+O\left(\frac{u}{\log^c u}\right)\qquad (u\to\infty)$$

holds for some constant $c \geqslant 5$. *Assume that* $w_x \ll x_3$, $\sqrt{x} \leqslant Y \leqslant x$ *and that* $1 \leqslant k \leqslant c_2 x_2$ *for some fixed positive constant* c_2. *Then, as* $x \to \infty$,

$$N_k(Y|w_x;\alpha) = (1+o(1))\frac{1}{d^k}N_k(Y|w_x).$$

Proof. This is a special case of Theorem 1.1 of De Koninck and Kátai [4]. ■

For each $n \in \mathbb{N}$, define

$$e(n) := \prod_{\substack{p^\alpha \| n \\ p \leqslant w_x}} p^\alpha \qquad \text{and} \qquad M(n) := \prod_{\substack{p^\alpha \| n \\ p > w_x}} p^\alpha.$$

Lemma 4.2. *Assume that the conditions of Lemma 4.1 are met. Set*

$$\begin{aligned} S_k(Y|w_x) &:= \#\{n = e(n)M(n) \leqslant Y : \omega(M(n)) = k\}, \\ S_k(Y|w_x;\alpha) &:= \#\{n = e(n)M(n) \leqslant Y : \omega(M(n)) = k,\ R(M(n)) = \alpha\}. \end{aligned}$$

Then, as $x \to \infty$,

$$S_k(Y|w_x;\alpha) = (1+o(1))\frac{1}{d^k}S_k(Y|w_x).$$

Proof. To prove Lemma 4.2, it is sufficient to observe that

$$S_k(Y|W_x;\alpha) = \sum_{\substack{\nu \leqslant x \\ p(\nu) \leqslant w_x}} N_k(\frac{Y}{\nu}|w_x;\alpha),$$

$$S_k(Y|W_x) = \sum_{\substack{\nu \leqslant x \\ p(\nu) \leqslant w_x}} N_k(\frac{Y}{\nu}|w_x),$$

and thereafter to apply Lemma 4.1 and sum over all $\nu \leqslant e^{w_x}$, say, and then show that the sum over those $\nu > e^{w_x}$ is negligible. ■

Lemma 4.3. *Let* $f(n)$ *be a real valued non negative arithmetic function. Let* a_n, $n = 1, \ldots, N$, *be a sequence of integers. Let* r *be a positive real number, and let* $p_1 < p_2 < \cdots < p_s \leqslant r$ *be prime numbers. Set* $Q = p_1 \cdots p_s$. *If* $d|Q$, *then let*

$$\sum_{\substack{n=1 \\ a_n \equiv 0 \pmod{d}}}^{N} f(n) = \eta(d)X + R(N,d),$$

where X *and* R *are real numbers,* $X \geqslant 0$, *and* $\eta(d_1 d_2) = \eta(d_1)\eta(d_2)$ *whenever* d_1 *and* d_2 *are co-prime divisors of* Q.

Assume that for each prime p, $0 \leqslant \eta(p) < 1$. *Setting*

$$I(N,Q) := \sum_{\substack{n=1 \\ (a_n,Q)=1}}^{N} f(n),$$

then the estimate

$$I(N,Q) = \{1 + 2\theta_1 H\} X \prod_{p|Q} (1+\eta(p)) + 2\theta_2 \sum_{\substack{d|Q \\ d \leqslant z^3}} 3^{\omega(d)} |R(N,d)|$$

holds uniformly for $r \geqslant 2$, $\max(\log r, S) \leqslant \frac{1}{8}\log z$, *where* $|\theta_1| \leqslant 1$, $|\theta_2| \leqslant 1$, *and*

$$H = \exp\left(-\frac{\log z}{\log r}\left\{\log\left(\frac{\log z}{S}\right) - \log\log\left(\frac{\log z}{S}\right) - \frac{2S}{\log z}\right\}\right)$$

and

$$S = \sum_{p|Q} \frac{\eta(p)}{1-\eta(p)} \log p.$$

When these conditions are satisfied, there exists an absolute positive constant c *such that* $2H \leqslant c < 1$.

Proof. For a proof, see Lemma 2.1, page 79, in the book of Elliott [5]. ■

Lemma 4.4. *Let* $\pi_r(x) := \#\{n \leqslant x : \omega(n) = r\}$. *There exist positive absolute constants* c_3, c_4 *such that*

$$\pi_r(x) \leqslant c_3 \frac{x}{\log x} \frac{(\log\log x + c_4)^{r-1}}{(r-1)!} \qquad (x \geqslant 3).$$

Proof. For a proof, see Hardy and Ramanujan [7]. ■

Lemma 4.5. *There exists an absolute constant* $c_5 > 0$ *such that, uniformly for* $2 \leqslant y \leqslant x$,

$$\Psi(x,y) \leqslant c_5 \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\}.$$

Proof. For a proof, see the book of Tenenbaum [10]. ■

5. Proof of Theorems 3.1 and 3.2

Let $\lambda(\alpha)$ stand for the length of the word α over A_d. Let $\beta = b_1 \dots b_k \in A_d^k$ and $\omega^*(n) := \sum_{\substack{p|n \\ p \notin \mathcal{R}}} 1$, so that $\omega^*(n) = \lambda(R(n))$.

Since $\mathcal{R}$ is a finite set, it is clear that

$$T_N := \sum_{n \leqslant N} \omega^*(n) = N \log\log N + O(N) \qquad (N \to \infty). \tag{5.1}$$

Now, for each positive integer j, let $Y_j = 2^j$ and $\eta_j := R(2^j) \dots R(2^{j+1}-1)$, so that $\xi = 0, \eta_1\eta_2 \dots$

We shall say that β is a *subword* in the word α if there exist $\gamma_1, \gamma_2 \in A_d^*$ such that $\alpha = \gamma_1\beta\gamma_2$. Moreover, let $u_\beta(\alpha)$ stand for the number of occurrences of β as a subword in α.

It is clear that, for each positive integer j such that $Y_j < N$, we have

$$\sum_{n=Y_j}^{Y_{j+1}-1} u_\beta(R(n)) \leqslant u_\beta(\eta_j) \leqslant \sum_{n=Y_j}^{Y_{j+1}-1} u_\beta(R(n)) + (k+1)Y_j \tag{5.2}$$

and

$$\sum_{n=Y_j}^{N} u_\beta(R(n)) \leqslant u_\beta(R(Y_j)\dots R(N)) \leqslant \sum_{n=Y_j}^{N} u_\beta(R(n)) + (k+1)(N-Y_j+1). \tag{5.3}$$

Assume that $w_x \ll x_6$, let j be fixed and set $x = Y_j$. Then, for any integer $n \in [Y_j, Y_{j+1}]$, we clearly have

$$u_\beta(R(M(n))) \leqslant u_\beta(R(n)) \leqslant \omega(e_n) + k + u_\beta(R(M(n))).$$

Observe that

$$\sum_{n=Y_j}^{N} (\omega(e_n) + k) \leqslant (N - Y_j)(2k + \omega(2x)).$$

We shall now provide asymptotic estimates for

$$K_j := \sum_{n=Y_j}^{Y_{j+1}-1} u_\beta(R(M(n))) \qquad \text{and} \qquad K_{N,Y_j} := \sum_{n=Y_j}^{N} u_\beta(R(M(n))). \tag{5.4}$$

To do so, we shall first find an upper bound for the number of those integers $n \in [Y_j, Y_{j+1} - 1]$ for which $\omega(M(n)) \geqslant 2x_2$. In fact, we will prove that

$$\Sigma_0 := \sum_{\substack{Y_j \leqslant n < Y_{j+1} \\ \omega(M(n)) \geqslant 2x_2}} \omega(M(n)) = O(Y_j). \tag{5.5}$$

Indeed, it follows from Lemma 4.4 that

$$\pi_r(Y_j) \leqslant \frac{c_3 Y_j}{\log Y_j} \frac{(\log\log Y_j + c_4)^{r-1}}{(r-1)!},$$

so that

$$\Sigma_0 = \sum_{r=\lfloor 2x_2 \rfloor}^{\infty} r\pi_r(Y_j) \leqslant c_3 \sum_{r \geqslant 2x_2} \frac{rY_j}{\log Y_j} \frac{(\log\log Y_j + c_4)^{r-1}}{(r-1)!} \ll Y_j,$$

thereby establishing our claim (5.5).

With this result in mind, we now only need to consider those integers n for which $r = \omega(M(n)) \leqslant 2x_2$.

So let $\alpha = e_1 \dots e_r \in A_d^r$, with $r \leqslant 2x_2$.

From Lemma 4.2, we have

$$\begin{aligned} S_r(Y|w_x, \alpha) &= \#\{n \leqslant Y : \omega(M(n)) = r,\ R(M(n)) = \alpha\} \\ &= (1+o(1))\frac{1}{d^r} S_r(Y|w_x), \end{aligned}$$

so that

$$\begin{aligned} &S_r(Y_{j+1}-1|w_x, \alpha) - S_r(Y_j - 1|w_x, \alpha) \\ &\qquad = (1+o(1))\frac{1}{d^r}\left(S_r(Y_{j+1}-1|w_x) - S_r(Y_j-1|w_x)\right). \end{aligned}$$

Now,

$$\begin{aligned} &S_r(N|w_x, \alpha) - S_r(Y_j - 1|w_x, 1) \\ &\qquad = (1+o(1))\frac{1}{d^r}\left(S_r(N|w_x) - S_r(Y_j|w_x)\right) + o(1)\frac{1}{d^r} S_r(N|w_x). \end{aligned}$$

From these observations and in light of (5.5), it follows that, as $x \to \infty$,

$$K_j = (1+o(1)) \sum_{r \leqslant 2x_2} \frac{1}{d^r}\left(\sum_{\alpha \in A_d^r} u_\beta(\alpha)\right)\left(S_r(2Y_j|w_x) - S_r(Y_j|w_x)\right) + O(Y_j). \quad (5.6)$$

On the other hand, we clearly have that

$$\sum_{\alpha \in A_d^r} u_\beta(\alpha) = \begin{cases} 0 & \text{if } r < k, \\ (r-k+1)d^{r-k} & \text{if } r \geqslant k. \end{cases}$$

Substituting this in (5.6), it follows that, as $x \to \infty$,

$$K_j = (1+o(1)) \sum_{r=k}^{\lfloor 2x_2 \rfloor} \frac{r-k+1}{d^k}\left(S_r(2Y_j|w_x) - S_r(Y_j|w_x)\right) + O(Y_j). \quad (5.7)$$

Since the contribution to K_j of those numbers r for which $|r - x_2| \geqslant x_2^{3/4}$ is clearly $o(x_2 Y_j)$, estimate (5.7) becomes

$$K_j = (1+o(1))\frac{x_2}{d^k} \sum_{|r-x_2|<x_2^{3/4}} \left(S_r(2Y_j|w_x) - S_r(Y_j|w_x)\right) + o(x_2 Y_j) \qquad (x \to \infty). \quad (5.8)$$

On the other hand, one can easily establish that

$$\begin{aligned} \sum_{|r-x_2|<x_2^{3/4}} \left(S_r(2Y_j|w_x) - S_r(Y_j|w_x)\right) &= (1+o(1))(2Y_j - Y_j) \\ &= (1+o(1))Y_j \qquad (x \to \infty). \end{aligned} \quad (5.9)$$

Substituting (5.9) in (5.8), we obtain

$$K_j = (1+o(1))\frac{x_2}{d^k} Y_j \qquad (x \to \infty). \tag{5.10}$$

We also need to estimate K_{N,Y_j} (defined in (5.4)) in the case $Y_j < N < Y_{j+1}$.
Let $\delta_1, \delta_2, \ldots$ be a sequence of positive numbers which tends to 0 very slowly.
If $N_{Y_j} \geqslant \delta_j Y_j$, then, in light of Lemma 4.1 and proceeding as above, one can prove that

$$K_{N,Y_j} = (1+o(1))\frac{x_2}{d^k}(N - Y_j) \qquad (x \to \infty),$$

while if $N_{Y_j} < \delta_j Y_j$, then

$$K_{N,Y_j} = O(\delta_j Y_j \log\log N) \qquad (Y_j \to \infty).$$

Hence, in light of these observations and of (5.10), it follows from inequalities (5.2) and (5.3) that

$$u_\beta(\eta_j) = (1+o(1))(Y_{j+1} - Y_j)\frac{\log\log Y_j}{d^k} \qquad (Y_j \to \infty) \tag{5.11}$$

and that

$$u_\beta(R(Y_j)\ldots R(N)) = (1+o(1))(N-Y_j)\frac{\log\log Y_j}{d^k} + O(\delta_j Y_j \log\log Y_j) \qquad (Y_j \to \infty). \tag{5.12}$$

Now, consider the d-ary expansion of the number ξ, that is $\xi = 0, R(1)R(2)\ldots$. Let $\xi^{(M)}$ be the rational approximation of ξ up to the M-th digit, that is $\xi^{(M)} = 0, R(1)R(2)\ldots R(M)$. We would like to approximate $u_\beta(\xi^{(M)})$. Given a fixed positive integer M, let N be defined implicitly by

$$\lambda(R(1)\ldots R(N)) \leqslant M < \lambda(R(1)\ldots R(N+1)).$$

Hence, in light of (5.1), we have that

$$M = T_N + O(N) = N\log\log N + O(N) \qquad (N \to \infty).$$

We therefore have that, for $Y_j \leqslant N < Y_{j+1}$,

$$u_\beta(\xi^{(M)}) = u_\beta(R(1)\ldots R(Y_j - 1)) + u_\beta(R(Y_j)\ldots R(N)) + O(\delta_j N \log\log N),$$

so that

$$\frac{u_\beta(\xi^{(M)})}{M} = \frac{u_\beta(R(1)\ldots R(Y_j-1))}{M} + \frac{u_\beta(R(Y_j)\ldots R(N))}{M} + O\left(\frac{\delta_j N \log\log N}{M}\right). \tag{5.13}$$

Taking into account estimates (5.11) and (5.12), it follows from (5.13) that

$$\begin{aligned}\frac{u_\beta(\xi^{(M)})}{M} &= (1+o(1))\frac{1}{d^k}\frac{T_{Y_j}}{M} + (1+o(1))\frac{T_N - T_{Y_j}}{d^k M}\\ &\quad + O\left(\frac{\delta_j N\log\log N}{M}\right) \qquad (N \to \infty),\end{aligned}$$

which implies, since $\delta_j \to 0$ as $j \to \infty$, that

$$\lim_{M\to\infty} \frac{u_\beta(\xi^{(M)})}{M} = \frac{1}{d^k},$$

thus completing the proof of Theorem 3.1.

The proof of Theorem 3.2 is very similar to that of Theorem 3.1 and, in fact, follows essentially from Theorem 3.3, which we shall now prove.

6. The proof of Theorem 3.3

Fix the integers K and a, and set $\Gamma := \gcd(K, a)$. If $\Gamma > 1$, then $p + a = K\mu$ implies that $p|\Gamma$, so that $\Delta_K(x) \leqslant \omega(a)$.

Hence, let us assume that $\Gamma = 1$. Let $q_1 < \cdots < q_M \leqslant x$ be those prime numbers for which $q_j + a \equiv 0 \pmod K$.

Let x be a large number and set $M = \pi(x; K, -a)$. Further define

$$Q(y) := \prod_{p<y} p \qquad \text{and} \qquad Q := Q(x^{\varepsilon_x}).$$

Choosing $r = x^{\varepsilon_x}$ and applying Lemma 4.3, we obtain that

$$I(M, Q) = \#\left\{q_j \leqslant x : \gcd\left(\frac{q_j + a}{K}, Q\right) = 1,\ j = 1, \dots, M\right\}.$$

For each squarefree number d, we have

$$\sum_{\substack{q_j \leqslant x \\ \frac{q_j+a}{K} \equiv 0 \pmod d}} 1 = \pi(x; Kd, -a) = \eta(d)\pi(x; K, -a) + R(x, d),$$

so that

$$R(x, d) = \pi(x; Kd, -a) - \eta(d)\pi(x; K, -a),$$

where η is a multiplicative function (whose domain is the squarefree numbers) defined on primes p by

$$\eta(p) = \begin{cases} 1/p & \text{if } p|K,\ \gcd(a, p) = 1, \\ 1/(p-1) & \text{if } \gcd(Ka, p) = 1, \\ 0 & \text{if } p|a. \end{cases}$$

Setting

$$E(x; d, \ell) := \pi(x; d, \ell) - \frac{\mathrm{li}(x)}{\varphi(d)},$$

it follows that

$$R(x, d) = E(x; d, \ell) - \eta(d)E(x; K, -a),$$

so that

$$|R(x,d)| \leqslant |E(x;d,\ell)| + \eta(d)|E(x;K,-a)|. \tag{6.1}$$

Observe that, if a is odd, then K is odd only when $p=2$, while if a is even, then $\Delta_{2+a}(x)=1$ for all $x \geqslant 2$, and $\Delta_K(x)=0$ for all other even K.

Let $\kappa(K)$ be defined as in (3.2).

Choose $z = x^{\delta_x}$, where δ_x is a function of x satisfying $\delta_x \to 0$ and $\delta_x/\varepsilon_x \to \infty$ as $x \to \infty$.

We have

$$S = \sum_{p|Q} \frac{\eta(p)}{1-\eta(p)} \log p = \sum_{p|Q} \eta(p)\log p + O\left(\sum_{p|Q} \eta^2(p)\log p\right).$$

It follows from this that there exists an absolute constant c_6 such that

$$S = \varepsilon_x \log x + c_6 + C(K) + o(1) \qquad (x \to \infty),$$

where $C(K)$ is a constant depending on K and satisfying

$$C(K) = O(1)\sum_{p|K} \frac{\log p}{p^2}.$$

Then, using Lemma 4.3, with H satisfying

$$H \leqslant \exp\left\{-\frac{1}{2}\frac{\delta_x}{\varepsilon_x}\log\frac{\delta_x}{\varepsilon_x}\right\},$$

we get

$$\Delta_K(x) = I(N,Q) = \kappa(K)\pi(x;K,-a)(1+2\theta_1 H) + 2\theta_2 \sum_{\substack{d\leqslant z^3\\ d|Q}} 3^{\omega(d)}|R(x,d)|,$$

from which it follows, using the Prime Number Theorem for primes in arithmetic progressions, that

$$\begin{aligned}\left|\Delta_K(x) - \kappa(K)\frac{\mathrm{li}(x)}{\varphi(K)}\right| &\leqslant \kappa(K)|E(x;K,-a)| \\ &\quad + 2\theta_1\kappa(K)\pi(x;K,-a)\cdot\exp\left\{-\frac{1}{2}\frac{\delta_x}{\varepsilon_x}\log\frac{\delta_x}{\varepsilon_x}\right\} \\ &\quad + 2\theta_2 \sum_{\substack{d\leqslant z^3\\ d|Q}} 3^{\omega(d)}\left\{|E(x;Kd,-a)| + \eta(d)|E(x;k,-a)|\right\}.\end{aligned} \tag{6.2}$$

Let us write $\kappa(K)$ in the form

$$\kappa(K) = \prod_{\substack{p|Q\\ p\neq 2}}\left(1-\frac{1}{p-1}\right)\rho(K),$$

where

$$\rho(K)=\prod_{\substack{p\mid K\\ p\neq 2}}\frac{1-1/p}{1-1/(p-1)}.$$

On the other hand, observe that by Mertens' Theorem, we have that, as $x\to\infty$,

$$\begin{aligned}\prod_{\substack{p\mid Q\\ p\neq 2}}\left(1-\frac{1}{p-1}\right)&=\exp\{-\log\log x^{\varepsilon_x}+c_7+o(1)\} \qquad (6.3)\\ &=(1+o(1))e^{c_7}(\log x^{\varepsilon_x})^{-1}.\end{aligned}$$

Summing the left hand side of (6.2) over $K\leqslant x^{\delta_x}$, $K\in\mathcal{K}_x$, we get

$$\sum_{\substack{K\leqslant x^{\delta_x}\\ K\in\mathcal{K}_x}}\left|\Delta_K(x)-\kappa(K)\frac{\mathrm{li}(x)}{\varphi(K)}\right|\leqslant\Sigma_1+\Sigma_2+\Sigma_3,$$

where

$$\Sigma_1\leqslant\sum_{K\leqslant x^{\delta_x}}\frac{c\,\mathrm{li}(x)}{\varphi(K)}\cdot\frac{1}{\varepsilon_x\log x}\exp\left\{-\frac{1}{2}\frac{\delta_x}{\varepsilon_x}\log\frac{\delta_x}{\varepsilon_x}\right\}, \qquad (6.4)$$

$$\Sigma_2\leqslant 2\theta_2\sum_{K\leqslant x^{\delta_x}}|E(x;K,-a)|\sum_{d\mid Q}\eta(d), \qquad (6.5)$$

$$\Sigma_3\leqslant\theta_3\sum_{\substack{M\leqslant x^{4\delta_x}\\ P(M)<x^{\varepsilon_x}}}|E(x;M,-a)|\sum_{\substack{d\leqslant z^3\\ d\mid Q,\ d\mid M}}3^{\omega(d)}, \qquad (6.6)$$

where we used (6.3) and (6.1).

On the other hand, one can prove that

$$\sum_{K\leqslant x^{\delta_x}}\frac{1}{\varphi(K)}\leqslant\prod_{p<x^{\varepsilon_x}}\left(1+\frac{1}{p-1}+\frac{1}{p(p-1)}+\ldots\right)\ll\log x^{\varepsilon_x}, \qquad (6.7)$$

$$\sum_{d\mid Q}\eta(d)\leqslant\prod_{p\mid Q}\left(1+\frac{1}{p-1}\right)\ll\log x^{\varepsilon_x}, \qquad (6.8)$$

$$\sum_{K\leqslant x^{\delta_x}}|E(x;K,-a)|\ll\frac{x}{\log^A x}, \qquad (6.9)$$

where A is an arbitrary large constant. Note that estimate (6.9) is a weak version of a theorem of Barban which is much weaker than the Bombieri-Vinogradov Theorem.

Using (6.7) in (6.4) as well as (6.8) and (6.9) in (6.5) yields the first two terms on the right hand side of (3.3). Hence, in order to complete the proof of Theorem 3.3, it will be sufficient to show that

$$\Sigma_3\ll\frac{x}{\log^A x}+\varepsilon_x\pi(x). \qquad (6.10)$$

First observe that

$$\sum_{\substack{d\leqslant z^3\\ d|Q,\ d|M}} 3^{\omega(d)} \leqslant 4^{\omega(M)}.$$

Using this and (6.6), we may write

$$\begin{aligned}\Sigma_3 &\ll \sum_{\substack{M\leqslant x^{4\delta_x}\\ P(M)<x^{\varepsilon_x}\\ \omega(M)\leqslant 30x_2}} |E(x;M,-a)|4^{\omega(M)} + \sum_{\substack{M\leqslant x^{4\delta_x}\\ P(M)<x^{\varepsilon_x}\\ \omega(M)>30x_2}} |E(x;M,-a)|4^{\omega(M)} \\ &= \Sigma_{3A}+\Sigma_{3B},\end{aligned} \tag{6.11}$$

say. Since it is clear that $|E(x;M,-a)| \leqslant \dfrac{c\,\mathrm{li}(x)}{\varphi(M)}$, it follows that

$$\begin{aligned}\Sigma_{3B} &\ll \mathrm{li}(x) \sum_{\substack{M\leqslant x^{4\delta_x}\\ P(M)<x^{\varepsilon_x}\\ \omega(M)>30x_2}} \frac{4^{\omega(M)}}{\varphi(M)} \\ &\ll \mathrm{li}(x)4^{-30x_2} \sum_{P(M)<x^{\varepsilon_x}} \frac{4^{2\omega(M)}}{\varphi(M)} \\ &\ll \mathrm{li}(x)4^{-30x_2} \prod_{p<x^{\varepsilon_x}} \left(1+\frac{16}{p-1}+\frac{16}{p(p-1)}+\dots\right) \\ &= \mathrm{li}(x)4^{-30x_2} \prod_{p<x^{\varepsilon_x}} \left(1+\frac{16p}{(p-1)^2}\right) \\ &= \mathrm{li}(x)4^{-30x_2} \exp\left\{\sum_{p<x^{\varepsilon_x}} \log\left(1+\frac{16p}{(p-1)^2}\right)\right\} \\ &\ll \mathrm{li}(x)4^{-30x_2}\exp\{16\log\log x^{\varepsilon_x}\} = \mathrm{li}(x)4^{-30x_2}(\varepsilon_x\log x)^{16} \\ &= \mathrm{li}(x)(\log x)^{-30\log 4+16}\varepsilon_x^{16} \ll \frac{x}{\log^2 x},\end{aligned} \tag{6.12}$$

since $-30\log 4+16<-1$.

On the other hand,

$$\Sigma_{3A} \leqslant 4^{30x_2} \sum_{M<x^{4\delta_x}} |E(x;M,-a)| \ll \frac{x}{\log^A x}. \tag{6.13}$$

Using (6.12) and (6.13) in (6.11) proves (6.10), thereby completing the proof of (3.3).

It remains to prove (3.4). Let us choose $\delta_x=\varepsilon_x^{\lambda}$, where $0<\lambda<1$. We first establish that

$$T(x) := \sum_{\substack{x^{\delta_x}<K\leqslant x\\ K\in\mathcal{K}_x}} \frac{\kappa(K)}{\varphi(K)} = o(1) \qquad (x\to\infty). \tag{6.14}$$

It follows from (3.2) that

$$\kappa(K) \leqslant \frac{\varphi(K)}{K} \prod_{2<p<x^{\varepsilon_x}} \left(1 - \frac{1}{p-1}\right) g(K), \tag{6.15}$$

where g is a multiplicative function defined on prime powers p^a by

$$g(p^a) = g(p) = \frac{1}{1-1/(p-1)} = \frac{p-1}{p-2}.$$

Therefore, letting $g(n) = \sum_{\delta | n} h(\delta)$, the function δ is a multiplicative function itself defined on prime powers p^a by

$$h(p^a) = \begin{cases} 1/(p-2) & \text{if } a=1 \text{ and } p>2, \\ 0 & \text{if } a=1 \text{ and } p=2, \\ 0 & \text{if } a \geqslant 2, \end{cases}$$

and since

$$\prod_{2<p<x^{\varepsilon_x}} \left(1 - \frac{1}{p-1}\right) < \frac{c_8}{\varepsilon_x \log x},$$

it follows from (6.15) that

$$T(x) \leqslant \frac{c_8}{\varepsilon_x \log x} \sum_{\substack{x^{\delta_x} < K \leqslant x \\ K \in \mathcal{K}_x}} \frac{g(K)}{K} = \frac{c_8}{\varepsilon_x \log x} T_1(x), \tag{6.16}$$

say. We certainly have that

$$T_1(x) \leqslant \sum_{\substack{\delta \leqslant x^{\delta_x/2} \\ P(\delta) < x^{\varepsilon_x}}} \frac{h(\delta)}{\delta} \sum_{\substack{L > x^{\varepsilon_x/2} \\ L \in \mathcal{K}_x}} \frac{1}{L} + \sum_{\delta > x^{\delta_x/2}} \frac{h(\delta)}{\delta} \sum_{P(L) < x^{\varepsilon_x}} \frac{1}{L} = T_2(x) + T_3(x), \tag{6.17}$$

say. In order to estimate $Z(x) := \sum_{\substack{L > x^{\varepsilon_x/2} \\ L \in \mathcal{K}_x}} \frac{1}{L}$, we proceed as follows.

Setting $U_0 = x^{\delta_x/2}$ and $U_j = 2^j U_0$ for $j = 1, 2, \ldots$, it follows from Lemma 4.5 that

$$\begin{aligned} Z(x) &\leqslant \sum_{\substack{j \geqslant 0 \\ U_j \leqslant x}} \frac{\Psi(2U_j, x^{\varepsilon_x})}{U_j} \qquad (6.18) \\ &\leqslant c_5 \sum_{0 \leqslant j \leqslant \frac{\log x - \log U_0}{\log 2}} \exp\left\{ -\frac{1}{2} \frac{\log x^{\delta_x/2} + j \log 2}{\varepsilon_x \log x} \right\} \\ &\leqslant c_5 \exp\left\{ -\frac{1}{4} \frac{\delta_x}{\varepsilon_x} \right\} \sum_{0 \leqslant j \leqslant \frac{\log x - \log U_0}{\log 2}} \exp\left\{ -\frac{1}{2} \frac{j \log 2}{\varepsilon_x \log x} \right\} \\ &\leqslant c_5 \exp\left\{ -\frac{1}{4} \frac{\delta_x}{\varepsilon_x} \right\} \cdot c_9 \cdot \frac{1}{\varepsilon_x \log x}. \end{aligned}$$

Using this estimate of $Z(x)$, we obtain that

$$T_2(x) \leqslant Z(x) \prod_{p<x^{\delta_x/2}} \left(1+\frac{h(p)}{p}\right) \leqslant \frac{c_{10}}{\varepsilon_x \log x} \exp\left\{-\frac{1}{4}\frac{\delta_x}{\varepsilon_x}\right\}. \tag{6.19}$$

On the other hand, since one can easily prove that

$$\sum_{\delta>x^{\delta_x/2}} \frac{h(\delta)}{\delta} < \frac{1}{x^{\delta_x/3}},$$

say, we have that

$$T_3(x) \leqslant \frac{1}{x^{\delta_x/3}} \sum_{P(L)<x^{\varepsilon_x}} \frac{1}{L} \leqslant \frac{1}{x^{\delta_x/3}} \prod_{p<x^{\varepsilon_x}} \left(1-\frac{1}{p}\right)^{-1} \ll \frac{1}{x^{\delta_x/3}\varepsilon_x \log x}. \tag{6.20}$$

Using (6.19) and (6.20) in (6.17), and bringing this into (6.16), then (6.14) follows, as required.

In light of (6.14), in order to complete the proof of (3.4), we only need to prove that, given any fixed positive number $\beta < 1$,

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \sum_{\substack{x^{\delta_x}<K<x^\beta \\ P(K)<x^{\varepsilon_x}}} \Delta_K(x) = 0. \tag{6.21}$$

Indeed, assuming that (6.21) holds and observing that $K|p+a$ with $p \leqslant x$ and $K > x^\beta$ implies that $P(p+a) \leqslant x^{1-\beta}$, then since

$$\frac{1}{\pi(x)} \sum_{x^{\delta_x}<K<x} \Delta_K(x) \leqslant \frac{1}{\pi(x)} \sum_{\substack{x^{\delta_x}<K<x^\beta \\ P(K)<x^{\varepsilon_x}}} \Delta_K(x) + \frac{1}{\pi(x)} \#\{p \leqslant x : P(p+a) \leqslant x^{1-\beta}\},$$

it follows that

$$\begin{aligned}\limsup_{x\to\infty} \frac{1}{\pi(x)} \sum_{x^{\delta_x}<K<x} \Delta_K(x) &\leqslant \limsup_{x\to\infty} \frac{1}{\pi(x)} \sum_{\substack{x^{\delta_x}<K<x^\beta \\ P(K)<x^{\varepsilon_x}}} \Delta_K(x) \\ &\quad + \limsup_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leqslant x : P(p+a) \leqslant x^{1-\beta}\}, \\ &\leqslant 0 + c(\beta),\end{aligned}$$

where $\lim_{\beta\to 1} c(\beta) = 0$, thus completing the proof of (3.4).

Hence, it remains to prove (6.21).

So, letting $0 < \beta < 1$, we shall prove that

$$\Sigma_\beta^{(a)} := \sum_{\substack{x^{\delta_x}<K<x^\beta \\ P(K)<x^{\varepsilon_x}}} \Delta_K(x) = o(\pi(x)). \tag{6.22}$$

For $K < x^\beta$, we have

$$\begin{aligned}\Delta_K(x) &\leqslant \frac{cx}{\varphi(K)\log(x/K)} \prod_{\substack{2<p<x^{\varepsilon_x}\\ \gcd(p,K)=1}} \left(1-\frac{1}{p-1}\right)\\ &\leqslant \frac{cx}{\varphi(K)(1-\beta)\log x}\cdot(\varepsilon_x\log x)^{-1}g(K).\end{aligned}$$

This implies that

$$\Sigma_\beta^{(a)} \leqslant \frac{c\,\mathrm{li}(x)}{(1-\beta)\varepsilon_x}\cdot\frac{1}{\log x}\sum_{x^{\delta_x}<K<x^\beta}\frac{g(K)}{\varphi(K)} = \frac{c\,\mathrm{li}(x)}{(1-\beta)\varepsilon_x}\cdot\frac{1}{\log x}D(x), \tag{6.23}$$

say. Set $t(K) := \dfrac{Kg(K)}{\varphi(K)}$ and observe that both $K/\varphi(K)$ and $g(K)$ are strongly multiplicative functions, and that, for each $p > 2$, $t(p) = pg(p)/(p-1)$, $g(p) = (p-1)/(p-2)$, so that $t(p) = p/(p-2)$, while $t(2) = g(2) = 0$.

Now, write $t(K) = \sum_{d|K}\ell(d)$, where ℓ is a multiplicative function with $\ell(p^a) = 0$ if $a \geqslant 2$, $p > 2$, and $\ell(p) = 1 - 2/(p-2)$.

With these notations in mind, we get that

$$\begin{aligned}D(x) &= \sum_{\substack{x^{\delta_x}<K<x^\beta\\ K\in\mathcal{K}_x}}\frac{t(K)}{K} = \sum_{\substack{x^{\delta_x}<dL<x^\beta\\ dL\in\mathcal{K}_x}}\frac{\ell(d)|\mu(d)|}{dL} \\ &= \sum_{\substack{x^{\delta_x}<dL<x^\beta\\ dL\in\mathcal{K}_x,\ d\leqslant x^{\delta_x/2}}}\frac{\ell(d)|\mu(d)|}{dL} + \sum_{\substack{x^{\delta_x}<dL<x^\beta\\ dL\in\mathcal{K}_x,\ d>x^{\delta_x/2}}}\frac{\ell(d)|\mu(d)|}{dL}\\ &= D_1(x)+D_2(x),\end{aligned} \tag{6.24}$$

say.

Now, one easily obtains that

$$\begin{aligned}D_1(x) &\leqslant Z(x)\sum_{P(d)\leqslant x^{\varepsilon_x}}\frac{\ell(d)|\mu(d)|}{d} \leqslant Z(x)\prod_{2<p<x^{\varepsilon_x}}\left(1+\frac{1}{p-2}\right),\\ D_2(x) &\leqslant Z(x)\sum_{L<x^\beta}\frac{1}{L} \leqslant Z(x)\prod_{p<x^{\varepsilon_x}}\left(1-\frac{1}{p}\right)^{-1}.\end{aligned}$$

Substituting these two last estimates in (6.24) and using the estimate of $Z(x)$ obtained in (6.18), it follows that

$$D(x) \leqslant Z(x)(\log x)\varepsilon_x \ll \exp\left\{-\frac{1}{4}\frac{\delta_x}{\varepsilon_x}\right\}.$$

Substituting this last estimate in (6.23) immediately yields (6.22), thus completing the proof of Theorem 3.3.

7. The proof of Theorems 3.4 and 3.5

Let $A_k = \{0, 1, \ldots, k-1\}$, $A_k^r = \underbrace{A_k \times \ldots \times A_k}_{r \text{ times}}$, $A_k^* = \cup_{r=0}^{\infty} A_k^r$ and $A_k^0 = \{\Lambda\}$.

Let $\eta_M \in A_k^M$ be the sequence of the first M digits in the expansion (3.5). Let $\alpha = b_1 \ldots b_d$ be an arbitrary word. Let $\nu(\theta)$ be the number of occurrences of α in the word θ, i.e. the number of those $\beta_1, \beta_2 \in A_k^*$ for which $\theta = \beta_1 \alpha \beta_2$. We shall prove that $\lim\limits_{M\to\infty} \dfrac{\nu(\eta_M)}{M} = \dfrac{1}{k^d}$, thus allowing us to complete the proof of Theorem 3.4.

Again, letting $\lambda(\alpha)$ stand for the length of the word α, we then have

$$\sum_{j=0}^{k-1} \omega(n+j) - k^2 \leqslant \lambda(S(E(n))) \leqslant \sum_{j=0}^{k-1} \omega(n+j).$$

Let $N = N(M)$ be defined as the largest integer for which

$$\lambda(S(E(1)) \ldots S(E(N))) \leqslant M < \lambda(S(E(1)) \ldots S(E(N)) S(E(N+1))).$$

With such a choice N, we clearly have

$$N \log\log N + O(N) = M.$$

Moreover,

$$0 \leqslant M - \lambda(S(E(1)) \ldots S(E(N))) \leqslant \sum_{\ell=0}^{k-1} \omega(N+\ell+1) = O(k \log N),$$

so that

$$(0 \leqslant) \nu_M(\alpha) - \nu_{M_1}(\alpha) = o(M) \qquad (M \to \infty).$$

Let $N = N(M)$ be fixed and consider the number

$$\eta(M_1) = S(E(1)) \ldots S(E(N)).$$

Moreover, let

$$\begin{aligned} \wp_1 &= \{p : p \leqslant k-1\}, \\ \wp_2 &= \{p : k \leqslant p \leqslant N^{1/\sqrt{\log\log N}}\}, \\ \wp_3 &= \{p : p > N^{1/\sqrt{\log\log N}}\}, \end{aligned}$$

and write each integer n as

$$n = A(n) B(n) C(n),$$

where $A(n) \in \mathcal{N}(\wp_1)$, $B(n) \in \mathcal{N}(\wp_2)$ and $C(n) \in \mathcal{N}(\wp_3)$.

We now proceed to estimate $\eta(M_1)$.

First observe that

$$\sum_{j=1}^{N} \nu(S(E(j))) \leqslant \nu(\eta_M) \leqslant \sum_{j=1}^{N} \nu(S(E(j))) + (k+1)N. \tag{7.1}$$

On the other hand,

$$\begin{aligned}\nu(S(A(E(n))B(E(n)))) \leqslant \nu(S(E(n))) \leqslant \nu(S(A(E(n))B(E(n)))) \\ + k + \underbrace{\nu(S(C(E(n))))}_{\leqslant k\sqrt{\log\log N}}.\end{aligned} \tag{7.2}$$

In light of (7.1) and (7.2), we have

$$\nu(\eta_M) = \sum_{n\leqslant N} \nu(S(A(E(n))B(E(n)))) + o(M) \qquad (M \to \infty). \tag{7.3}$$

We shall now see that we can ignore all those positive integers $n \leqslant N$ for which one of the following conditions hold:

(a) $\max\limits_{0\leqslant j\leqslant k-1} A(n+j) > A_N$,

(b) $B(E(n)) > N^{A_N/\sqrt{\log\log N}}$,

(c) $|\omega(E(n)) - k\log\log N| > \dfrac{1}{A_N}\log\log N$,

where A_N is a function which tends to $+\infty$ slowly as $N \to \infty$, but with $A_N = o(\sqrt{\log\log N})$.

Let $\mathcal{T}_N$ be the set of those integers $n \leqslant N$ which do not satisfy any of the above conditions. Then, one can show that

$$\sum_{\substack{n\leqslant N\\ n\notin\mathcal{T}_N}} \nu(S(A(E(n))B(E(n)))) = o(M) \qquad (M \to \infty). \tag{7.4}$$

Hence, it follows from (7.3) and (7.4) that

$$\begin{aligned}\nu(\eta_M) &= \sum_{n\in\mathcal{T}_N} \nu(S(A(E(n))B(E(n)))) + o(M) \\ &= \sum_{n\in\mathcal{T}_N} \nu(S(B(E(n)))) + o(M).\end{aligned} \tag{7.5}$$

Let R run over the integers belonging to $\mathcal{N}(\wp_2)$ and not exceeding $N^{A_N/\sqrt{\log\log N}}$. Assume that $m_0, \ldots, m_{k-1} \in \mathcal{N}(\wp_2)$ are co-prime integers such that $R = m_0 \ldots m_{k-1}$, and consider the set

$$\mathcal{U}(m_0, \ldots, m_{k-1}) := \#\{n \in \mathcal{T}_N : \frac{n+\ell}{m_\ell} \in \mathcal{N}(\wp_1 \cup \wp_3),\ \ell = 0, \ldots, k-1\} \tag{7.6}$$

and assume that there exist integers $\xi_0, \dots, \xi_{k-1} \in \mathcal{N}(\wp_1)$ which satisfy the property that $\xi_j \| n+j$ $(j = 0, \dots, k-1)$. Then, define $\xi^* = \mathrm{LCM}[\xi_0, \dots, \xi_{k-1}]$ and set $\xi^{**} = \xi^* \prod_{q \in \wp_1} q$.

Let $r_0, \dots, r_{k-1}$ be representatives of all the k residue classes mod ξ^{**} which satisfy

$$n \equiv r_j \pmod{\xi^{**}} \Longrightarrow \xi_j \| n+j \qquad (j = 0, \dots, k-1).$$

Let $\mathcal{U}_{r,\xi^{**}}(m_0, \dots, m_{k-1})$ be the subset of those integers $n \in \mathcal{U}(m_0, \dots, m_{k-1})$ which satisfy the additional condition $n \equiv r (\mathrm{mod}\ \xi^{**})$. Then, further define

$$V(R) := \bigcup_{\substack{m_0 \cdots m_{k-1} = R \\ \gcd(m_i, m_j) = 1}} \mathcal{U}(m_0, \dots, m_{k-1}),$$

$$V_{r \mod \xi^{**}}(R) := \bigcup_{\substack{m_0 \cdots m_{k-1} = R \\ \gcd(m_i, m_j) = 1}} \mathcal{U}_{r \bmod \xi^{**}}(m_0, \dots, m_{k-1}).$$

Let us now evaluate the size of the set $\mathcal{U}_{r \bmod \xi^{**}}(m_0, \dots, m_{k-1})$.

Clearly, $n \equiv r \mod \xi^{**}$ means that there exists an integer s such that $n = r + s\xi^{**}$. Now, the condition $n + \ell \equiv 0 \mod m_\ell$, for $\ell = 0, \dots, k-1$, determines exactly one $s_0 \mod R$ for which both the congruence $\ell + r + s_0\xi^{**} \equiv 0 \mod m_\ell$ and the condition $\xi_\ell \| \ell + r + s_0\xi^{**}$ hold. Now, let $n = r + s_0\xi^{**} + t\xi^{**}R$ and define

$$G_\ell(t) := \frac{r + s_0\xi^{**}}{\xi_\ell m_\ell} + \frac{\xi^{**}R}{\xi_\ell m_\ell}\, t,$$
$$G(t) := G_0(t) \cdots G_{k-1}(t).$$

Clearly, $\gcd(G_\ell(t), \wp_1) = 1$. We will count the number of positive integers $t \leqslant N/(\xi^{**}R)$ for which $G(t) \in \mathcal{N}(\wp_3)$, that is for which $p(G(t)) > N^{1/\sqrt{\log\log N}}$. In order to estimate the number of these t's, one can use the Fundamental Lemma (see Theorem 2.6 in Halberstam and Richert [6]), so that, assuming ξ^{**} is not too large, say $\xi^{**} \leqslant \log\log N$, then we obtain

$$\#\mathcal{U}_{r \bmod \xi^{**}}(m_0, \dots, m_{k-1}) = (1 + o(1)) \frac{N}{\xi^{**}R} \prod_{k \leqslant p \leqslant N^{1/\sqrt{\log\log N}}} \left(1 - \frac{k}{p}\right) \tag{7.7}$$
$$= (1 + o(1)) \frac{N}{\xi^{**}R} \rho_N \qquad (N \to \infty),$$

say. Clearly, we have that

$$\log \rho_N = -k \log\log N^{1/\sqrt{\log\log N}} + c + o(1) \qquad (N \to \infty).$$

Substituting this estimate in (7.7), we see that the resulting asymptotic estimate depends only on $R = m_0 \cdots m_{k-1}$.

Now, the number of possible factorizations of R as $m_0 \cdots m_{k-1}$, with $\gcd(m_i, m_j) = 1$ when $0 \leqslant i < j \leqslant k-1$, is $k^{\omega(R)}$. So, write R as $R = q_1^{e_1} \cdots q_v^{e_v}$, where $q_1 < \cdots < q_v$ and each $e_i \in \mathbb{N}$. Then, let $\varepsilon_1 \dots \varepsilon_v \in A_k^v$ be an arbitrary word, and further define $m_\ell = \prod_{\substack{\varepsilon_\alpha = \ell \\ \alpha = 1, \dots, v}} q_\alpha^{e_\alpha}$.

To complete the proof of Theorem 3.4, we shall need the following proposition.

Proposition 7.1. *Let d and k be fixed positive integers. For each integer $v \geqslant d$, consider the words $\varepsilon_1 \dots \varepsilon_v \in A_k^v$. Given a subword $\beta = b_1 \dots b_d$, let $\sigma_\beta(\varepsilon_1 \dots \varepsilon_v)$ stand for the number of occurrences of the subword β in the word $\varepsilon_1 \dots \varepsilon_v$. Then, there exists a constant $c_9 > 0$ such that*

$$\frac{1}{k^v} \sum_{\varepsilon_1 \dots \varepsilon_v \in A_k^v} \left(\sigma_\beta(\varepsilon_1 \dots \varepsilon_v) - \frac{v}{k^d} \right)^2 \leqslant c_9 v.$$

Proof. First observe that, since the word $\beta = b_1 \dots b_d$ occupies d positions, it leaves out $v-d$ free positions, implying that

$$\sum_{\varepsilon_1 \dots \varepsilon_v \in A_k^v} \sigma_\beta(\varepsilon_1 \dots \varepsilon_v) = k^{v-d}(v-d+1). \tag{7.8}$$

On the other hand,

$$\begin{aligned} \sum_{\varepsilon_1 \dots \varepsilon_v \in A_k^v} \sigma_\beta^2(\varepsilon_1 \dots \varepsilon_v) &= 2 \sum_{v_1 \leqslant v_2 - d} \#\{\varepsilon_j \in A_k : \varepsilon_{v_1} \dots \varepsilon_{v_1+d-1} = \beta \\ &\qquad\qquad \text{and } \varepsilon_{v_2} \dots \varepsilon_{v_2+d-1} = \beta\} + O(vk^{v-d}) \\ &= 2 \sum_{v_1 \leqslant v_2 - d} k^{v-2d} + O(vk^{v-d}) \\ &= k^{v-2d} \left(2 \frac{(v-d)^2}{2} + O(v) \right) \\ &= v^2 k^{v-2d} + O(vk^{v-2d}). \end{aligned} \tag{7.9}$$

Calling upon (7.8) and (7.9), we obtain that

$$\begin{aligned} \sum_{\varepsilon_1 \dots \varepsilon_v \in A_k^v} \left(\sigma_\beta(\varepsilon_1 \dots \varepsilon_v) - \frac{v}{k^d} \right)^2 &= \sum_{\varepsilon_1 \dots \varepsilon_v \in A_k^v} \sigma_\beta^2(\varepsilon_1 \dots \varepsilon_v) \\ &\quad - 2 \frac{v}{k^d} \sum_{\varepsilon_1 \dots \varepsilon_v \in A_k^v} \sigma_\beta(\varepsilon_1 \dots \varepsilon_v) + k^v \frac{v^2}{k^{2d}} \\ &= v^2 k^{v-2d} + O(vk^{v-2d}) - 2 \frac{v}{k^d} k^{v-d}(v-d+1) \\ &\quad + v^2 k^{v-2d} \\ &= k^{v-2d} \cdot (2v(d-1) + O(v)). \end{aligned}$$

Dividing both sides of the above relation by k^v yields the desired estimate. ■

We are now ready to complete the proof of Theorem 3.4.

With the definition of $\mathcal{U}_{r_j (\text{mod } \xi^{**})}(m_0, \dots, m_{k-1})$ given in (7.6), one can write relation (7.5) as

$$\nu(\eta_M) = \sum_{R \in \mathcal{N}(\wp_2)} \sum_{\xi^{**} < \log\log N} \sum_{r_0, \dots, r_{k-1}} \sum_{R = m_0 \dots m_{k-1}} F_j(\alpha) + o(M), \tag{7.10}$$

where $F_j(\alpha)$ stands for the number of occurrences of α in the set $\mathcal{U}_{r_j(\bmod\ \xi^{**})}(m_0,\dots,m_{k-1})$.

Now, for every partition of $R=m_0\dots m_{k-1}$ with $\gcd(m_i,m_j)=1$ for $0\leqslant i<j\leqslant k-1$, there is a unique word $\varepsilon_1\dots\varepsilon_v\in A_k^v$ such that the number of occurrences of α in $\varepsilon_1\dots\varepsilon_v$ is equal to $\nu_\alpha(S(B(E(n))))$ for $n\in\mathcal{U}_{r_j(\bmod\ \xi^{**})}(m_0,\dots,m_{k-1})$.

But it follows from Proposition 1 that

$$\nu_\alpha(S(B(E(n))))=\frac{\omega(R)}{k^d}+o(1)\omega(R)$$

if $n\in\mathcal{U}_{r_j(\bmod\ \xi^{**})}(m_0,\dots,m_{k-1})$ with the exception of no more than $o(k^v)$ choices of $m_0,\dots,m_{k-1}$ satisfying $R=m_0\dots m_{k-1}$. Hence, it follows from (7.10) that

$$\nu(\eta_M)=\sum_{R\in\mathcal{N}(\wp_2)}\sum_{\xi^{**}<\log\log N}\sum_{r_0,\dots,r_{k-1}}\frac{\omega(R)}{q^d}\#\mathcal{U}_{r_j\ (\bmod\ \xi^{**})}(R)+o(M). \tag{7.11}$$

Replacing $\omega(R)$ in the above by $\log\log N$, then (7.11) becomes

$$\begin{aligned}\nu(\eta_M)&=(1+o(1))\frac{\log\log N}{k^d}\sum_{R\in\mathcal{N}(\wp_2)}\sum_{\xi^{**}<\log\log N}\sum_{r_0,\dots,r_{k-1}}\#\mathcal{U}_{r_j\ (\bmod\ \xi^{**})}(R)+o(M)\\&=(1+o(1))\frac{\log\log N}{k^d}(N+o(N))+o(M)\qquad(N\to\infty).\end{aligned}$$

Dividing both sides of the above formula by M yields the result

$$\lim_{M\to\infty}\frac{\nu(\eta_M)}{M}=\frac{1}{k^d},$$

which completes the proof of Theorem 3.4.

The proof of Theorem 3.5 can be obtained along the same lines.

References

[1] D.G. Champernowne, *The construction of decimals normal in the scale of ten*, J. London Math. Soc. **8** (1933), 254-260.

[2] A.H. Copeland and P. Erdős, *Note on normal numbers*, Bull. Amer. Math. Soc. **52** (1946), 857-860.

[3] H. Davenport and P. Erdős, *Note on normal decimals*, Can. J. Math. **4** (1952), 58-63.

[4] J.M. De Koninck and I. Kátai, *On the distribution of subsets of primes in the prime factorization of integers*, Acta Arithmetica **72**(2) (1995), 169-200.

[5] P.D.T.A. Elliott, *Probabilistic Number Theory I, Mean Value Theorems*, Springer-Verlag, Berlin, 1979

[6] H.H. Halberstam and H.E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[7] G.H.Hardy and S.Ramanujan, *The normal number of prime factors of a number n*, Quart.J.Math. **48** (1917), 76-92.

[8] M.G. Madritsch, J.M. Thuswaldner and R.F. Tichy, *Normality of numbers generated by the values of entire functions*, J. of Number Theory **128** (2008), 1127-1145.

[9] Y. Nakai and I. Shiokawa, *Normality of numbers generated by the values of polynomials at primes*, Acta Arith. **81**(4) (1997), 345-356.

[10] G. Tenenbaum, *Introduction à la théorie analytique des nombres*, Collection Échelles, Belin, 2008.

Addresses: Jean-Marie De Koninck: Dép. de mathématiques, Université Laval, Québec, Québec G1V 0A6, Canada;
Imre Kátai: Computer Algebra Department, Eötvös Loránd University, 1117 Budapest, Pázmány Péter Sétány I/C, Hungary.

E-mail: jmdk@mat.ulaval.ca, katai@compalg.inf.elte.hu

Received: 22 July 2010