Normal numbers generated using the smallest prime factor function

Jean-Marie De Koninck[1] and Imre Kátai[2]

Édition du 13 mai 2014

Abstract

In a series of papers, we constructed large families of normal numbers using the distribution of the values of the largest prime factor function. Here, letting $p(n)$ stand for the smallest prime factor of n, we show how a concatenation of the successive values of $p(n)$ can yield a normal number in any given basis $q \geq 2$. We further expand on this idea to create various large families of normal numbers.

AMS Subject Classification numbers: 11K16, 11N37, 11N41
Key words: normal numbers, smallest prime factor

1 Introduction

Given an integer $q \geq 2$, we say that an irrational number η is a *q-normal number* if the q-ary expansion of η is such that any preassigned sequence of length $k \geq 1$, taken within this expansion, occurs with the expected limiting frequency, namely $1/q^k$.

In a series of recent papers, we constructed large families of normal numbers using the distribution of the values of the largest prime factor function (see for instance [3], [4] and [5]).

Here, letting $p(n)$ stand for the smallest prime factor of n, we show how a concatenation of the successive values of $p(n)$ can yield a normal number. We further expand on this idea to create various large families of normal numbers.

2 Notation

Let $\wp$ stand for the set of all the prime numbers. The letters p and π with or without subscript will always denote prime numbers. The letter c, with or without subscript, always denotes a positive constant, but not necessarily the same at each occurrence. At times, we will use the notation $x_1 = \log x$, $x_2 = \log\log x$, $x_3 = \log\log\log x$.

We let $P(n)$ stand for the largest prime factor of the integer $n \geq 2$, with $P(1) = 1$, and we let $\omega(n)$ stand for the number of distinct prime factors of the integer $n \geq 2$,

[1]Research supported in part by a grant from NSERC.
[2]Research supported by ELTE IK.

with $\omega(1) = 0$. As usual, φ will denote the Euler totient function and $\mathrm{li}(x) := \int_2^x \frac{dt}{\log t}$ stands for the logarithmic integral. Finally, we set $\pi(x; k, \ell) := \#\{p \le x : p \equiv \ell \pmod{k}$.

Let $q \ge 2$ be a fixed integer and let $A = A_q = \{0, 1, 2, \ldots, q-1\}$. Given an integer $t \ge 1$, an expression of the form $i_1 i_2 \ldots i_t$, where each $i_j \in A_q$, is called a *word* of length t. Given a word α, we shall write $\lambda(\alpha) = t$ to indicate that α is a *word* of length t. We shall also use the symbol Λ to denote the *empty word*.

Then, $A^t = A_q^t$ will stand for the set of words of length t over A, while A^* will stand for the set of all words over A regardless of their length, including the empty word Λ. Observe that the concatenation of two words $\alpha, \beta \in A^*$, written $\alpha\beta$, also belongs to A^*. Finally, given a word α and a subword β of α, we will denote by $F_\beta(\alpha)$ the number of occurrences of β in α, that is, the number of pairs of words μ_1, μ_2 such that $\mu_1 \beta \mu_2 = \alpha$.

Given a positive integer n, we write its q-ary expansion as

$$n = \varepsilon_0(n) + \varepsilon_1(n) q + \cdots + \varepsilon_t(n) q^t,$$

where $\varepsilon_i(n) \in A$ for $0 \le i \le t$ and $\varepsilon_t(n) \ne 0$. To this representation, we associate the word

$$\overline{n} = \varepsilon_0(n)\varepsilon_1(n)\ldots\varepsilon_t(n) \in A^{t+1}.$$

For convenience, if $n \le 0$, we will write $\overline{n} = \Lambda$, the empty word.

The number of digits of such a number $\overline{n}$ will be $\lambda(\overline{n}) = \left\lfloor \dfrac{\log n}{\log q} \right\rfloor + 1$.

Finally, given a sequence of integers $a(1), a(2), a(3), \ldots$, we will say that the concatenation of their q-ary digit expansions $\overline{a(1)}\,\overline{a(2)}\,\overline{a(3)} \ldots$, denoted by $\mathrm{Concat}(\overline{a(n)} : n \in \mathbb{N})$, is a *normal sequence* if the number $0.\overline{a(1)}\,\overline{a(2)}\,\overline{a(3)}\ldots$ is a q-normal number.

3 Main results

Theorem 1. *The expression* $n_1 = \mathit{Concat}(\overline{p(n)} : n \in \mathbb{N})$ *is a normal sequence.*

Theorem 2. *Let* $R \in \mathbb{Z}[x]$ *be a polynomial such that* $R(x) > 0$ *for all* $x > 0$ *and satisfying* $\lim_{x\to\infty} R(x) = \infty$. *The expression* $n_2 = \mathit{Concat}(\overline{R(p(n))} : n \in \mathbb{N})$ *is a normal sequence.*

Theorem 3. *Let* $a \in \mathbb{Z}$ *be an even integer. The expression* $n_3 = \mathit{Concat}(\overline{p(\pi + a)} : \pi \in \wp)$ *is a normal sequence.*

Remark 1. *Observe that the particular case* $a = 0$ *has been proved by Davenport and Erdős [2].*

Theorem 4. *Let* $a \in \mathbb{Z}$ *be an even integer and let* R *be as in Theorem 2. The expression* $n_4 = \mathit{Concat}(\overline{R(p(\pi + a))} : \pi \in \wp)$ *is a normal sequence.*

We will only provide the proofs of Theorems 1 and 3, since the proofs of Theorems 2 and 4 can be obtained along the same lines.

4 Preliminary results

For each integer $n \geq 2$, let $L(n) = \left\lceil \frac{\log n}{\log q} \right\rceil$.

The next two lemmas follow as a particular case of Theorem 1 of Bassily and Kátai [1].

Lemma 1. *Let κ_u be a function of u such that $\kappa_u > 1$ for all u. Given a word $\beta \in A_q^k$ and setting*

$$V_\beta(u) := \#\left\{p \in \wp : u \leq p \leq 2u \text{ such that } \left|F_\beta(\overline{p}) - \frac{L(u)}{q^k}\right| > \kappa_u \sqrt{L(u)}\right\},$$

then, there exists a positive constant c such that

$$V_\beta(u) \leq \frac{cu}{(\log u)\kappa_u^2}.$$

Lemma 2. *Let κ_u be as in Lemma 1. Given $\beta_1, \beta_2 \in A_q^k$ with $\beta_1 \neq \beta_2$, set*

$$\Delta_{\beta_1,\beta_2}(u) := \#\left\{p \in \wp : u \leq p \leq 2u \text{ such that } |F_{\beta_1}(\overline{p}) - F_{\beta_2}(\overline{p})| > \kappa_u \sqrt{L(u)}\right\}.$$

Then, for some positive constant c,

$$\Delta_{\beta_1,\beta_2}(u) \leq \frac{cu}{(\log u)\kappa_u^2}.$$

Lemma 3. *Let $f(n)$ be a non negative real valued arithmetic function. Let a_n, $n = 1, \ldots, N$, be a sequence of integers. Let r be a positive real number, and let $p_1 < p_2 < \cdots < p_s \leq r$ be prime numbers. Set $Q = p_1 \cdots p_s$. If $d|Q$, then let*

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

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

Assume that for each prime p, $0 \leq \kappa(p) < 1$. Setting

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

then the estimate

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

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

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

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

Proof. This result is Lemma 2.1 in the book of Elliott [6]. □

5 Proof of Theorem 1

Let x be a large number, but fixed. Consider the interval

$$I_x := \left[\left\lfloor \frac{x}{2} \right\rfloor + 1, \lfloor x \rfloor\right)$$

and the following two truncated words of n_1:

$$\eta_x := \text{Concat}(\overline{p(n)} : n \leq x), \qquad \rho_x := \text{Concat}(\overline{p(n)} : n \in I_x).$$

Let β be an arbitrary word in A_q^k.

Letting ℓ_0 be the largest integer such that $2^{\ell_0} < x$, it is clear that

$$F_\beta(\eta_x) = \sum_{\ell=0}^{\ell_0} F_\beta(\rho_{x/2^\ell}) + O(\log x), \tag{5.1}$$

$$F_\beta(\rho_{x/2^\ell}) = \sum_{n \in I_{x/2^\ell}} F_\beta(\overline{p(n)}) + O\left(\frac{x}{2^\ell}\right), \tag{5.2}$$

where the error term on the right hand side of (5.1) accounts for the cases where the word β overlaps two consecutive intervals $I_{x/2^{\ell+1}}$ and $I_{x/2^\ell}$. Note that here and throughout this section, the constants implied by the Landau notation $O(\cdots)$ may depend of the particular basis q and on the particular word β.

Hence, in light of (5.1) and (5.2), in order to prove that n_1 is a normal sequence, it will be sufficient to show that, given any two words $\beta_1, \beta_2 \in A_q^k$, we have

$$\frac{|F_{\beta_1}(\rho_x) - F_{\beta_2}(\rho_x)|}{\lambda(\rho_x)} \to 0 \qquad \text{as } x \to \infty. \tag{5.3}$$

We first start by establishing the exact order of $\lambda(\rho_x)$.

For each $Q \in \wp$, we let

$$\nu_x(Q) = \#\{n \in I_x : p(n) = Q\}.$$

Let ε_x be a function such that $\lim_{x\to\infty} \varepsilon_x = 0$. Let also $Y_x < Z_x$ be two positive functions tending to infinity with x, that we will specify later. It is clear, using Mertens' formula, that, as $x \to \infty$,

$$\nu_x(Q) = (1+o(1))\frac{x}{2Q}\prod_{\substack{\pi<Q\\ \pi\in\wp}}\left(1-\frac{1}{\pi}\right) = (1+o(1))\frac{e^{-\gamma}}{2}\frac{x}{Q\log Q} \tag{5.4}$$

uniformly for $Y_x < Q \le x^{\varepsilon_x}$ (here γ stands for the Euler-Mascheroni constant). By a sieve approach, we may say that for some absolute constant $c_1 > 0$, we have

$$\nu_x(Q) \begin{cases} \le c_1 \frac{x}{Q\log Q} & \text{for all } Q \le \sqrt{x}, \\ \le \frac{x}{Q} & \text{for } \sqrt{x} < Q \le x. \end{cases} \tag{5.5}$$

We may then write

$$\begin{aligned} \lambda(\rho_x) &= \sum_{Q<Y_x} \nu_x(Q)\lambda(\overline{Q}) + \sum_{Y_x\le Q<Z_x} \nu_x(Q)\lambda(\overline{Q}) + \sum_{Z_x\le Q\le x} \nu_x(Q)\lambda(\overline{Q}) + O(x) \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3 + O(x), \end{aligned} \tag{5.6}$$

say. As we will see, the main contribution will come from the term Σ_2.

Using (5.4) and (5.5), we easily obtain

$$\Sigma_1 \le c_2 x \sum_{Q<Y_x} \frac{1}{Q\log Q}\cdot \log Q \le c_3 x \log\log Y_x, \tag{5.7}$$

$$\Sigma_3 \le c_4 x \sum_{Z_x\le Q\le x} \frac{1}{Q} \le c_5 x \log\left(\frac{\log x}{\log Z_x}\right). \tag{5.8}$$

Choosing Y_x so that $\log Y_x = (\log x)^{\varepsilon_x}$ and Z_x so that $\dfrac{\log x}{\log Z_x} = (\log x)^{\varepsilon_x}$, it follows from (5.7) and (5.8) that, as $x \to \infty$,

$$\Sigma_1 = o(x\log\log x), \tag{5.9}$$

$$\Sigma_3 = o(x\log\log x). \tag{5.10}$$

Now, in light of (5.4), we have, as $x \to \infty$,

$$\begin{aligned} \Sigma_2 &= \sum_{Y_x\le Q<Z_x} \nu_x(Q)\lambda(\overline{Q}) \\ &= (1+o(1))c_6 x \sum_{Y_x\le Q<Z_x} \frac{\lambda(\overline{Q})}{Q\log Q} = (1+o(1))\frac{c_7 x}{\log q}\sum_{Y_x\le Q<Z_x}\frac{1}{Q} \end{aligned}$$

$$(5.11) \quad = \quad (1+o(1))c_7 x \log\left(\frac{\log Z_x}{\log Y_x}\right) = (1+o(1))c_7 x \log\log x + O(x\varepsilon_x \log\log x),$$

for some positive constants c_6 and c_7.

Hence, gathering estimates (5.9), (5.10) and (5.11) and substituting them into (5.6), we obtain that

$$\lambda(\rho_x) = c_7 x \log\log x + o(x \log\log x),$$

thus establishing that the true order of $\lambda(\rho_x)$ is $x\log\log x$. Therefore, in light of our ultimate goal (5.3), we now only need to show that

$$|F_{\beta_1}(\rho_x) - F_{\beta_2}(\rho_x)| = o(x\log\log x) \qquad (x \to \infty). \tag{5.12}$$

To accomplish this, using the same approach as above, we easily get that

$$|F_{\beta_1}(\rho_x) - F_{\beta_2}(\rho_x)| \le \sum_{Y_x < Q < Z_x} \left|F_{\beta_1}(\overline{Q}) - F_{\beta_2}(\overline{Q})\right| \nu_x(Q) + o(x\log\log x). \tag{5.13}$$

We further set ℓ_1 as the largest integer such that $2^{\ell_1+1} \le Y_x$ and ℓ_2 as the smallest integer such that $2^{\ell_2+1} \ge Z_x$. We then write the interval $[Y_x, Z_x]$ as a subset of the union of a finite number of intervals, namely as follows:

$$[Y_x, Z_x] \subseteq \bigcup_{\ell=\ell_1}^{\ell_2} \left[\frac{x}{2^{\ell+1}}, \frac{x}{2^{\ell}}\right], \tag{5.14}$$

that is the union of a finite number of intervals of the form $[u, 2u]$.

For each of these intervals $[u, 2u]$, we have

$$T(u) := \sum_{u \le Q \le 2u} \left|F_{\beta_1}(\overline{Q}) - F_{\beta_2}(\overline{Q})\right| \nu_x(Q) = S_1(u) + S_2(u), \tag{5.15}$$

where $S_1(u)$ is the same as $T(u)$ but with the restriction that the sum runs only over those primes $Q \in [u, 2u]$ for which

$$\left|F_{\beta_1}(\overline{Q}) - F_{\beta_2}(\overline{Q})\right| \le \kappa_u \sqrt{L(u)},$$

while $S_2(u)$ accounts for the other primes $Q \in [u, 2u]$, namely those for which

$$\left|F_{\beta_1}(\overline{Q}) - F_{\beta_2}(\overline{Q})\right| > \kappa_u \sqrt{L(u)}.$$

Using Lemma 2 and (5.5), we thus have that, for some positive constants c_8 and c_9,

$$S_1(u) \quad \le \quad c_8 \sum_{u \le Q \le 2u} \kappa_u \sqrt{\log u}\; \nu_x(Q) \le c_8 \kappa_u \sqrt{\log u}\; x \sum_{u \le Q \le 2u} \frac{1}{Q\log Q}$$

$$(5.16) \qquad \leq \quad c_9\frac{x\kappa_u}{(\log u)^{3/2}}.$$

On the other hand, using the trivial estimate $F_{\beta_i}(\overline{Q}) \leq \lambda(\overline{Q}) \ll \log u$, we easily get, again using Lemma 2 and (5.5), that, for some positive constant c_{10},

$$(5.17) \qquad S_2(u) \leq \frac{c_{10}x}{u}\frac{u}{(\log u)\kappa_u^2} = \frac{c_{10}x}{(\log u)\kappa_u^2}.$$

Substituting (5.16) and (5.17) in (5.15), we obtain that

$$(5.18) \qquad T(u) \leq cx\left(\frac{\kappa_u}{(\log u)^{3/2}} + \frac{1}{(\log u)\cdot\kappa_u^2}\right).$$

We now choose $\kappa_u = \log\log\log x$. Then, in light of (5.14) and using (5.18), we may conclude that

$$\sum_{Y_x<Q<Z_x} |F_{\beta_1}(\rho_x) - F_{\beta_2}(\rho_x)| \leq \sum_{\ell=\ell_1}^{\ell_2} T\left(\frac{x}{2^\ell}\right) \leq o(x\log\log x),$$

which in light of (5.13) proves (5.12), thereby completing the proof of Theorem 1.

6 Proof of Theorem 3

We let x be a large number and turn our attention to the truncated word

$$\sigma_x = \mathrm{Concat}(\overline{p(\pi+a)} : \pi \in I_x),$$

of which we first plan to estimate the length $\lambda(\sigma_x)$.

For each prime number U, let

$$M_x(U) = \#\{\pi \in I_x : p(\pi+a) = U\}.$$

This allows us to write

$$(6.1) \qquad \lambda(\sigma_x) = \sum_{U\in\wp} M_x(U)\lambda(\overline{U}) = \sum_{\substack{U<x^{\varepsilon_x}\\ U\in\wp}} + \sum_{\substack{U\geq x^{\varepsilon_x}\\ U\in\wp}} = \Sigma_1 + \Sigma_2,$$

say. Using Theorem 4.2 of Halberstam and Richert [7], we get that

$$\begin{aligned} \Sigma_2 &\leq (\log x)\cdot\#\{\pi < x : p(\pi+a) \geq x^{\varepsilon_x}\} \\ (6.2) \qquad &\leq c\frac{x\log x}{\log x}\prod_{p<x^{\varepsilon_x}}\left(1-\frac{1}{p}\right) \leq c_1\frac{x}{\varepsilon_x \log x}, \end{aligned}$$

by Mertens' estimate.

Let us choose ε_x so that $1/\varepsilon_x$ tends monotonically to infinity, but very slowly. We will now use Lemma 3 and the Bombieri-Vinogradov theorem to estimate $M_x(U)$ for $U < x^{\varepsilon_x}$ for almost all U. Choose $\kappa_U = 1/\sqrt{\varepsilon_U}$.

Following the notation of Lemma 3, we have

$$T_U = \prod_{p<U} p, \quad p_1 < \cdots < p_s(\leq U), \quad \Delta = \pi(U) - 1,$$

$$\left(\frac{x}{2} \leq\right) \pi_1 < \cdots < \pi_N(\leq x), \quad \pi_j + a \equiv 0 \pmod{U},$$

$$a_n = \pi_n + a \text{ for } n = 1, 2, \ldots, N, \quad f(n) = 1 \text{ for all } n \in \mathbb{N}.$$

Moreover, for each $d|T_U$,

$$\pi(I_x; dU, -a) = \sum_{a_n \equiv 0 \pmod{d}} f(n) = \frac{1}{\varphi(d)(U-1)} (\mathrm{li}(x) - \mathrm{li}(x/2)) + R(N, dU, -a),$$

say. We have

$$|R(N, dU, -a)| \leq \left|\pi(x; dU, -a) - \frac{\mathrm{li}(x)}{\varphi(dQ)}\right| + \left|\pi(\frac{x}{2}; dU, -a) - \frac{\mathrm{li}(x/2)}{\varphi(dQ)}\right|.$$

Let η be the multiplicative function defined on the squarefree integers by

$$\eta(p) = \begin{cases} 1/(p-1) & \text{if } p \nmid a, \\ 0 & \text{if } p \mid a. \end{cases}$$

We then have

$$S = \sum_{\substack{p|T_U \\ p\nmid a}} \frac{\log p}{p-2} = \log U + O(1).$$

Then, the condition

$$\frac{1}{8} \log z \geq \max(\log \pi(U), \log U)$$

clearly holds for every large U. Further set

$$H = H_U = \exp\left\{-\kappa_U \left(\log \kappa_U - \log\log \kappa_U - \frac{2}{\kappa_U}\right)\right\}.$$

We then have

$$M_x(U) = \{1 + 2\theta_1 H\} \frac{\mathrm{li}(x) - \mathrm{li}(x/2)}{U-1} \prod_{\substack{2<p<U \\ p\nmid a}} \left(1 - \frac{1}{p-1}\right) + B(U), \tag{6.3}$$

where

$$B(U) = 2\theta_2 \sum_{\substack{d|T_U \\ d \leq U^{\kappa_U}}} 3^{\omega(d)} |R(N, d)|,$$

and where $|\theta_1| \le 1$, $|\theta_2| \le 1$.

On the one hand, there exists a constant $A_1 = A_1(a) > 0$ such that

$$\prod_{\substack{2<p<U \\ p\nmid a}} \left(1 - \frac{1}{p-1}\right) = (1+o(1))\frac{A_1}{\log U} \qquad (U \to \infty). \tag{6.4}$$

On the other hand,

$$\begin{aligned}
\sum_{U \le x^{\varepsilon_x}} B(U) \;&\le\; \sum_{U \le x^{\varepsilon_x}} \sum_{\substack{d|T_U \\ d \le U^{\kappa_U}}} 3^{\omega(d)} \left| \pi(x; dU, -a) - \frac{\mathrm{li}(x)}{\varphi(dU)} \right| \\
&\quad + \sum_{U \le x^{\varepsilon_x}} \sum_{\substack{d|T_U \\ d \le U^{\kappa_U}}} 3^{\omega(d)} \left| \pi(x/2; dU, -a) - \frac{\mathrm{li}(x/2)}{\varphi(dU)} \right| \\
&= S_1(x) + S_2(x), \qquad (6.5)
\end{aligned}$$

say. We have $dU \le U^{\kappa_U+1}$. Set $m = dU$. Since $U = P(m)$, it follows that m determines d and U uniquely.

We shall now provide an estimate for $S_1(x)$ by using the Brun-Titchmarsh inequality and the Bombieri-Vinogradov inequality. So, let $B > 0$ and $E > 0$ be arbitrary numbers. We then have

$$\begin{aligned}
S_1(x) \;&\ll\; \sum_{\substack{m \le x^{\sqrt{\varepsilon_x}+\varepsilon_x} \\ \omega(m) \le Bx_2}} 3^{Bx_2} \left| \pi(x; m, -a) - \frac{\mathrm{li}(x)}{\varphi(m)} \right| + \sum_{\substack{m \le x^{\sqrt{\varepsilon_x}+\varepsilon_x} \\ \omega(m) > Bx_2}} 3^{\omega(m)} \frac{\mathrm{li}(x)}{\varphi(m)} \\
&\ll \frac{x \cdot 3^{Bx_2}}{x_1^E} + \frac{\mathrm{li}(x)}{3^{Bx_2}} \sum_{\substack{m \le x^{1/4} \\ P(m) < x^{\varepsilon_x}}} \frac{3^{2\omega(m)}}{\varphi(m)} \\
&\ll \frac{x \cdot 3^{Bx_2}}{x_1^E} + \frac{\mathrm{li}(x)}{3^{Bx_2}} \prod_{p < x^{\varepsilon_x}} \left(1 + \frac{9p}{(p-1)^2}\right). \qquad (6.6)
\end{aligned}$$

It follows from (6.6) that, given any fixed number $A > 0$, an appropriate choice of B and E will lead to

$$S_1(x) \ll \frac{\mathrm{li}(x)}{\log^A x}. \tag{6.7}$$

Proceeding in a similar manner, we easily obtain that

$$S_2(x) \ll \frac{\mathrm{li}(x)}{\log^A x}. \tag{6.8}$$

Using (6.7) and (6.8) in (6.5), and combining this with (6.4) and (6.3) in our estimate (6.2), and recalling (6.1), we obtain

$$\lambda(\sigma_x) = \sum_{U \in \wp} M_x(U)\lambda(\overline{U}) = \Sigma_1 + \Sigma_2 \ll \Sigma_1 + \frac{x}{(\log x)\varepsilon_x}. \tag{6.9}$$

Let us now write

$$\Sigma_1 = \sum_{U<\log x} + \sum_{\log x \le U < x^{\varepsilon_x}} = T_1 + T_2, \tag{6.10}$$

say.

First observe that, using (6.3), as $x \to \infty$,

$$\begin{aligned} T_1 &= \sum_{U<\log x} (1+o(1)) \frac{A_1(\mathrm{li}(x) - \mathrm{li}(x/2))}{(U-1)\log U} \lambda(\overline{U}) + O\left(\frac{\mathrm{li}(x)}{\log^A x}\right) \\ &\ll \mathrm{li}(x) \sum_{U<\log x} \frac{1}{U} + O\left(\frac{\mathrm{li}(x)}{\log^A x}\right) \\ &\ll \mathrm{li}(x) \cdot x_3, \end{aligned} \tag{6.11}$$

while

$$\begin{aligned} T_2 &\ll (\mathrm{li}(x) - \mathrm{li}(x/2)) \sum_{\log x \le U < x^{\varepsilon_x}} \frac{1}{(U-1)\log U} \left\lceil \frac{\log U}{\log q} \right\rceil \\ &\ll \frac{1}{\log q} (\mathrm{li}(x) - \mathrm{li}(x/2)) \sum_{\log x \le U < x^{\varepsilon_x}} \frac{1}{U} \\ &= (1+o(1)) \frac{1}{\log q} (\mathrm{li}(x) - \mathrm{li}(x/2)) \log\log x. \end{aligned} \tag{6.12}$$

Gathering (6.11), (6.12) and (6.10) in (6.9), we get

$$\lambda(\sigma_x) \ll \frac{1}{2\log q \cdot \log x} (\mathrm{li}(x) - \mathrm{li}(x/2))\, x_2 \ll \frac{x x_2}{\log x}. \tag{6.13}$$

Let $\beta_1, \beta_2 \in A_q^k$ and set $\Delta(\alpha) = F_{\beta_1}(\alpha) - F_{\beta_2}(\alpha)$. We will prove that

$$\lim_{x\to\infty} \frac{|\Delta(\sigma_x)|}{\lambda(\sigma_x)} = 0. \tag{6.14}$$

First, observe that it is clear that

$$|\Delta(\sigma_x)| \le \sum_{U\in\wp} M_x(U) |\Delta(\overline{U})| + O(1) \sum_{U\in\wp} M_x(U).$$

By using (6.2), we obtain that

$$\sum_{U > x^{\varepsilon_x}} M_x(U) \le c \frac{x}{\varepsilon_x \log^2 x}.$$

By using (6.7) and (6.8), we obtain that

$$\sum_{\substack{U\in\wp \\ U \le x^{\varepsilon_x}}} B(U) |\Delta(\overline{U})| \le \log x \cdot \sum_{\substack{U\in\wp \\ U \le x^{\varepsilon_x}}} B(U) \le \frac{x}{\log^2 x},$$

provided $x > x_0$.

Thus, by using (6.3) and (6.11), we obtain that

$$|\Delta(\sigma_x)| \le \sum_{\substack{U\in\wp \\ \log x \le U \le x^{\varepsilon_x}}} c\frac{x}{\log x}\cdot\frac{\Delta(\overline{U})}{U\log U} + O\left(\frac{x\cdot x_3}{\log x}\right).$$

By using Lemma 2, it follows that

$$\sum_{\substack{U\in\wp \\ V\le U\le 2V}} |\Delta(\overline{U})| \le \frac{cV\log V}{\log V\cdot\kappa_V^2} + \frac{cV}{\log V}\cdot\kappa_V\cdot\log V = \frac{cV}{\kappa_V^2} + cV\kappa_V.$$

Thus,

$$\sum_{\substack{U\in\wp \\ V\le U\le 2V}} \frac{|\Delta(\overline{U})|}{U\log U} \le \frac{c}{\log V\cdot\kappa_V^2} + \frac{c\kappa_V}{\log^{3/2} V}. \tag{6.15}$$

Let us apply this with $V = V_j$ for $j = 0, 1, \ldots, j_0$, where $V_0 = \log x$, $V_j = 2^j V_0$, with $V_{j_0} \le x^{\varepsilon_x} < V_{j_0+1}$.

Thus, it follows from (6.15) that

$$\begin{aligned}\sum_{\substack{U\in\wp \\ \log x \le U \le x^{\varepsilon_x}}} \frac{|\Delta(\overline{U})|}{U\log U} &\le \frac{c}{\kappa_{V_0}^2}\sum_{j=0}^{j_0}\frac{1}{\log(V_0\cdot 2^j)} + c\kappa_{V_{j_0+1}}\sum_{j=0}^{j_0}\frac{1}{\log^{3/2} V_j} \\ &= W_1 + W_2, \end{aligned} \tag{6.16}$$

say. Since

$$W_1 \le \frac{c_1}{\kappa_{V_0}^2}\log j_0 \le \frac{c_1 x_2}{\kappa_{V_0}^2}$$

and noting that $\kappa_{V_0} \to \infty$ as $x \to \infty$, and since

$$W_2 \le c\kappa_x \sum_{j\ge 0}\frac{1}{(\log V_0 + j)^{3/2}} \le \frac{c_2\kappa_x}{x_2^{1/2}},$$

it follows from (6.16), that if we choose $\kappa_x \le \sqrt{x_2}$ say, then

$$\sum_{\substack{U\in\wp \\ \log x \le U \le x^{\varepsilon_x}}} \frac{|\Delta(\overline{U})|}{U\log U} = o(x_2),$$

which, in light of (6.13), proves (6.14) and thus completes the proof of Theorem 3.

7 Further remarks

Using the same approach, one can also prove the following two theorems.

Theorem 5. *Let* $G(n) = n^2 + 1$ *and set*

$$\begin{aligned} \xi_1 &= Concat(\overline{p(G(n))} : n \in \mathbb{N}), \\ \xi_2 &= Concat(\overline{p(G(\pi))} : \pi \in \wp). \end{aligned}$$

Then ξ_1 *and* ξ_2 *are q-normal sequences.*

We further let $p_k(n)$ stand for the k-th smallest prime factor of n, that is, if $n = q_1^{\alpha_1} \cdots q_r^{\alpha_r}$, where $q_1 < \cdots < q_r$ are primes and each α_i an integer, then

$$p_k(n) = \begin{cases} q_k & \text{if } k \le r, \\ 1 & \text{if } k > r. \end{cases}$$

Theorem 6. *Let* $G(x) = a_r x^r + a_{r-1} x^{r-1} + \cdots + a_0 \in \mathbb{Z}[x]$ *be irreducible and satisfying* $(a_r, a_{r-1}, \ldots, a_0) = 1$, $a_r > 0$ *and* $G(x) > 0$ *for* $x > x_0$. *Then*

$$\eta_k = Concat\left(\overline{p_k(G(n))} : x_0 < n \in \mathbb{N}\right)$$

is a q-normal sequence.

Observe that the proof of Theorem 6 is very similar to that of Theorem 1. Indeed, we first define

$$\kappa_x := \text{Concat}\left(\overline{p_k(G(n))} : n \in I_x\right),$$

where $I_x = [\lfloor x/2 \rfloor + 1, \lfloor x \rfloor)$. Then, for each prime Q, we set

$$T(Q) := \#\{n \in I_x : p_k(G(n)) = Q\},$$

so that

$$\lambda(\kappa_x) = \sum_{n \in I_x} \lambda(\overline{p_k(G(n))}) = \sum_{Q \le x} \lambda(\overline{Q}) T(Q).$$

As can be shown using sieve methods, the main contribution to the above sum comes from those primes $Q \le x^{1/2k}$, while that coming from the primes $Q > x^{1/2k}$ can be neglected. This allows us to establish that the order of $\lambda(\kappa_x)$ is $x(\log\log x)^k$.

Then, it is enough to prove that, given an arbitrary $t \in \mathbb{N}$ and any two words $\beta_1, \beta_2 \in A_q^t$,

$$\frac{|F_{\beta_1}(\kappa_x) - F_{\beta_2}(\kappa_x)|}{\lambda(\kappa_x)} \to 0 \qquad \text{as } x \to \infty$$

and this is done by showing that

$$|F_{\beta_1}(\kappa_x) - F_{\beta_2}(\kappa_x)| = o(x(\log\log x)^k) \qquad \text{as } x \to \infty.$$

References

[1] N.L. Bassily and I. Kátai, *Distribution of consecutive digits in the q-ary expansions of some sequences of integers*, Journal of Mathematical Sciences **78**, no. 1 (1996), 11-17.

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

[3] J.M. De Koninck and I. Kátai, *On a problem on normal numbers raised by Igor Shparlinski*, Bulletin of the Australian Mathematical Society **84** (2011), 337–349.

[4] J.M. De Koninck and I. Kátai, *Some new methods for constructing normal numbers*, Annales des Sciences Mathématiques du Québec **36** (2012), no.2, 349-359.

[5] J.M. De Koninck and I. Kátai, *Construction of normal numbers using the distribution of the k-th largest prime factor*, Journal of the Australian Mathematical Society **88** (2013), 158-168.

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

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

Jean-Marie De Koninck
Dép. de mathématiques et de statistique
Université Laval
Québec
Québec G1V 0A6
Canada
jmdk@mat.ulaval.ca

Imre Kátai
Computer Algebra Department
Eötvös Loránd University
1117 Budapest
Pázmány Péter Sétány I/C
Hungary
katai@compalg.inf.elte.hu

JMDK, le 13 mai 2014; fichier: normal-smallest-prime-factor-2014.tex