COLLOQUIUM MATHEMATICUM

VOL. 135 2014 NO. 1

NORMAL NUMBERS AND THE MIDDLE PRIME FACTOR OF AN INTEGER

BY

JEAN-MARIE DE KONINCK (Québec) and IMRE KÁTAI (Budapest)

Abstract. Let $p_m(n)$ stand for the middle prime factor of the integer $n \geq 2$. We first establish that the size of $\log p_m(n)$ is close to $\sqrt{\log n}$ for almost all n. We then show how one can use the successive values of $p_m(n)$ to generate a normal number in any given base $D \geq 2$. Finally, we study the behavior of exponential sums involving the middle prime factor function.

1. Introduction. Given an integer $D \geq 2$, a *D-normal number* is an irrational number ξ such that any preassigned sequence of l digits occurs in the D-ary expansion of ξ at the expected frequency, namely $1/D^l$.

In a series of recent papers, we constructed large families of D-normal numbers using the distribution of the values of the largest prime factor function $P(n)$ (see for instance [2], [3] and [4]). We also showed [5] how one can use the large prime divisors of an integer to construct normal numbers. Recently, we proved [6] that the concatenation of the successive values of $p(n)$, the smallest prime factor of n, in a given base $D \geq 2$, yields a D-normal number.

Given an integer $n \geq 2$, write it as $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, where $p_1 < \cdots < p_k$ are its distinct prime factors and $\alpha_1, \ldots, \alpha_k$ are positive integers. We let $p_m(n) = p_{\max(1,\lfloor k/2 \rfloor)}$ and say that $p_m(n)$ is the "middle" prime factor of n. Recently, De Koninck and Luca [7] showed that as $x \to \infty$,

$$\sum_{n \leq x} \frac{1}{p_m(n)} = \frac{x}{\log x} \exp\left((1 + o(1))\sqrt{2 \log\log x \ \log\log\log x}\right),$$

thus answering in part a question raised by Paul Erdős.

Here, we first establish that the size of $\log p_m(n)$ is, for almost all n, close to $\sqrt{\log n}$, and then we show how one can use the middle prime factor of an integer to generate a normal number in any given base $D \geq 2$. Finally, we study the behavior of exponential sums involving the middle prime factor function.

2010 *Mathematics Subject Classification*: Primary 11K16; Secondary 11N37.
Key words and phrases: normal numbers, middle prime factor.

DOI: 10.4064/cm135-1-5

2. Notation. The letters p, q and π, with or without subscript, will always denote prime numbers. The letter c, with or without subscript, will always denote a positive constant, but not necessarily the same at each occurrence.

Let $D \geq 2$ be a fixed integer and let $A = A_D = \{0, 1, \ldots, D-1\}$. Given an integer $t \geq 1$, an expression of the form $i_1 \ldots i_t$, where each $i_j \in A_D$, 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*. For each $t \in \mathbb{N}$, we let $A^t = A_D^t$ stand for the set of words of length t over A, while $A^* = A_D^*$ 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 D-ary expansion as

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

where $\varepsilon_i(n) \in A$ for $0 \leq i \leq t$ and $\varepsilon_t(n) \neq 0$. 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 $n \leq 0$, we will write $\overline{n} = \Lambda$. Observe that the number of digits of such a number $\overline{n}$ will thus be $\lambda(\overline{n}) = \lfloor (\log n)/\log D \rfloor + 1$.

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

3. Main results

THEOREM 3.1. *Let $g(x)$ be a function which tends to infinity with x but arbitrarily slowly. Set $x_2 = \log\log x$. Then, as $x \to \infty$,*

$$\frac{1}{x}\#\left\{n \in [x, 2x] : e^{-\sqrt{x_2}\,g(x)} \leq \frac{\log p_m(n)}{\sqrt{\log x}} \leq e^{\sqrt{x_2}\,g(x)}\right\} \to 1, \tag{3.1}$$

$$\frac{1}{x}\#\left\{n \leq x : e^{-\sqrt{x_2}\,g(x)} \leq \frac{\log p_m(n)}{\sqrt{\log x}} \leq e^{\sqrt{x_2}\,g(x)}\right\} \to 1. \tag{3.2}$$

Analogously, as $x \to \infty$,

$$\frac{1}{x}\#\left\{n \leq x : \left|\log\log p_m(n) - \tfrac{1}{2}x_2\right| \leq \sqrt{x_2}\,g(x)\right\} \to 1. \tag{3.3}$$

THEOREM 3.2. *The sequence* $\mathrm{Concat}(\overline{p_m(n)} : n \in \mathbb{N})$ *is D-normal in every basis $D \geq 2$.*

From here on, we will be using the standard notation $e(y) := \exp(2\pi i y)$. We now introduce the sum

$$T(x) := \sum_{n \le x} \log p_m(n).$$

THEOREM 3.3. *Consider the real-valued polynomial* $Q(x) = \alpha_k x^k + \cdots + \alpha_1 x$, *where at least one of the coefficients* $\alpha_k, \ldots, \alpha_1$ *is irrational, and set*

$$E_Q(x) := \sum_{n \le x} \log p_m(n) \cdot e(Q(p_m(n))).$$

Then,

$$E_Q(x) = o(T(x)) \qquad (x \to \infty).$$

REMARK 3.4. Observe that Theorem 3.3 includes the interesting case $Q(x) = \alpha x$, where α is an arbitrary irrational number.

4. Preliminary results

LEMMA 4.1. *Given a positive integer* k, *let* β_1 *and* β_2 *be distinct words belonging to* A_D^k. *Let* $c_0 > 0$ *be an arbitrary number and consider the intervals*

$$J_w := [w, w + w/\log^{c_0} w] \qquad (w > 1).$$

Further, let $\pi(J_w)$ *stand for the number of prime numbers belonging to the interval* J_w. *Then*

$$\frac{1}{\pi(J_w)} \sum_{p \in J_w} \frac{|F_{\beta_1}(\overline{p}) - F_{\beta_2}(\overline{p})|}{\log p} \to 0 \qquad \text{as } w \to \infty.$$

Proof. This result is a consequence of Theorem 1 in the paper of Bassily and Kátai [1]. ∎

LEMMA 4.2. *Let*

$$E_x := \sum_{\substack{n \le x \\ q p_m(n) | n \\ p_m(n)/3 < q < 3 p_m(n)}} \log p_m(n).$$

Then there exists a positive constant c *such that*

$$E_x \le c x \log\log x.$$

Proof. We have

$$E_x \le \sum_{p \le x} \log p \sum_{\substack{qpr \le x \\ p/3 < q < 3p}} 1 \le x \sum_{p \le x} \frac{\log p}{p} \sum_{p/3 < q < 3p} \frac{1}{q}$$

$$\le c_1 x \sum_{p \le x} \frac{1}{p} \le c_2 x \log\log x. \ ∎$$

LEMMA 4.3. *Let $Q(x) = \alpha_k x^k + \cdots + \alpha_1 x$ be a real-valued polynomial such that at least one of its coefficients $\alpha_k, \ldots, \alpha_1$ is irrational. If $p_1 < p_2 < \cdots$ stands for the sequence of primes, then*

$$\sum_{n \le x} e(Q(p_n)) = o(x) \qquad \text{as } x \to \infty.$$

Proof. For a proof of this result, see Chapters 7 and 8 in the book of I. M. Vinogradov [8]. ∎

5. Proof of Theorem 3.1. Let

$$y = \exp\left(\sqrt{\log x}\right), \quad \text{so that} \quad \log\log y = \tfrac{1}{2}x_2. \tag{5.1}$$

Then set

$$\omega_y(n) = \sum_{\substack{p \mid n \\ p < y}} 1, \qquad R_y(n) = \sum_{\substack{p \mid n \\ p > y}} 1, \qquad \Delta_y(n) = \omega_y(n) - R_y(n).$$

It is well known that, if $\varepsilon_x \to 0$ arbitrarily slowly as $x \to \infty$, then

$$\frac{1}{x}\#\left\{n \le x : |\omega(n) - x_2| > \frac{1}{\varepsilon_x}\sqrt{x_2}\right\} \to 0 \qquad \text{as } x \to \infty.$$

On the other hand, from the Turán–Kubilius inequality and in light of our choice of y given by (5.1), we have

$$\sum_{n \le x} \left(\omega_y(n) - \tfrac{1}{2}x_2\right)^2 = \sum_{n \le x} |\omega_y(n) - \log\log y|^2 = O(x x_2).$$

Secondly,

$$\begin{aligned} \left|R_y(n) - \tfrac{1}{2}x_2\right|^2 &\le \left(|\omega(n) - x_2| + \left|\omega_y(n) - \tfrac{1}{2}x_2\right|\right)^2 \\ &\le 2\left((\omega(n) - x_2)^2 + \left(\omega_y(n) - \tfrac{1}{2}x_2\right)^2\right), \end{aligned} \tag{5.2}$$

where we used the basic inequality $(a+b)^2 \le 2(a^2 + b^2)$ valid for all real numbers a and b. Then, summing both sides of (5.2) for $n \le x$, we obtain, for some positive constant C,

$$\sum_{n \le x} |\Delta_y(n)|^2 \le \sum_{n \le x} 2\left|\omega_y(n) - \tfrac{1}{2}x_2\right|^2 + \sum_{n \le x} 2\left|R_y(n) - \tfrac{1}{2}x_2\right|^2 \le C x x_2. \tag{5.3}$$

It follows from (5.3) that

$$|\Delta_y(n)| \le \frac{1}{\varepsilon_x}\sqrt{x_2} \quad \text{for all but at most } o(x) \text{ integers } n \le x. \tag{5.4}$$

Let us now choose z and w so that

$$\log z = (\log y) e^{-\sqrt{x_2}\, g(x)}, \qquad \log w = (\log y) e^{\sqrt{x_2}\, g(x)}.$$

Since

$$\sum_{z<p<y} \frac{1}{p} = \log \frac{\log y}{\log z} + o(1) = \sqrt{x_2}\, g(x) + o(1) = A(x) + o(1),$$

say, and similarly,

$$\sum_{y<p<w} \frac{1}{p} = \log \frac{\log w}{\log y} + o(1) = \sqrt{x_2}\, g(x) + o(1) = A(x) + o(1),$$

setting

$$\omega_{[a,b]}(n) := \sum_{\substack{p|n \\ p\in[a,b]}} 1,$$

and again using the Turán–Kubilius inequality, we have

$$\sum_{n\le x} (\omega_{[z,y]}(n) - A(x))^2 \le CxA(x),$$
$$\sum_{n\le x} (\omega_{[y,w]}(n) - A(x))^2 \le CxA(x),$$

from which it follows that

$$|\omega_{[z,y]}(n) - A(x)| \le \frac{1}{\varepsilon_x}\sqrt{A(x)}, \tag{5.5}$$

$$|\omega_{[y,w]}(n) - A(x)| \le \frac{1}{\varepsilon_x}\sqrt{A(x)}. \tag{5.6}$$

Now, recall that from (5.4), we only need to consider those $n \le x$ for which

$$|\omega_y(n) - R_y(n)| \le \frac{1}{\varepsilon_x}\sqrt{x_2},$$

and for which (5.5) and (5.6) hold. So, let us choose $\varepsilon_x = 2/g(x)$, in which case we have $A(x) = \sqrt{x_2} \cdot g(x) = (2/\varepsilon_x)\sqrt{x_2}$. Thus, assuming first that $0 \le R_y(n) - \omega_y(n) < \frac{1}{\varepsilon_x}\sqrt{x_2}$, we have $p_m(n) > y$ and by (5.6), $p_m(n) < w$, provided x is large enough. On the other hand, if $-\frac{1}{\varepsilon_x}\sqrt{x_2} \le R_y(n) - \omega_y(n) \le 0$, then $p_m(n) \le y$ and by (5.5), $p_m(n) > z$, provided x is large enough. Hence, in any case, we get

$$z \le p_m(n) \le w,$$

which proves (3.2), from which (3.1) and (3.3) follow as well, thus completing the proof of Theorem 3.1.

6. Proof of Theorem 3.2. Let x be a fixed large number. Let $L_x := \{n \in \mathbb{N} : \lfloor x \rfloor \le n \le \lfloor 2x \rfloor - 1\}$ and set

$$\rho_x := \text{Concat}(\overline{p_m(n)} : n \in L_x).$$

It is clear that

$$\lambda(\rho_x) = \sum_{n\in L_x} \lambda(\overline{p_m(n)}), \tag{6.1}$$

$$F_\beta(\rho_x) = \sum_{n\in L_x} F_\beta(\overline{p_m(n)}) + O(x), \tag{6.2}$$

$$\lambda(\overline{p}) = \frac{\log p}{\log D} + O(1). \tag{6.3}$$

It follows from (6.1), (6.3) and Theorem 3.1 that there exists $c_1 > 0$ such that

$$\lambda(\rho_x) \geq c_1 x\sqrt{\log x}\exp(-\sqrt{x_2}g(x)). \tag{6.4}$$

Given arbitrary distinct words $\beta_1, \beta_2 \in A_D^k$, we set

$$\Delta(\alpha) := F_{\beta_1}(\alpha) - F_{\beta_2}(\alpha) \qquad (\alpha \in A_D^*).$$

Our main task will be to prove that

$$\lim_{x\to\infty} \frac{\Delta(\rho_x)}{\lambda(\rho_x)} = 0. \tag{6.5}$$

This will prove that, for any word $\beta \in A_D^k$,

$$\frac{F_\beta(\rho_x)}{\lambda(\rho_x)} - \frac{1}{D^k} = o(1) \qquad \text{as } x \to \infty, \tag{6.6}$$

and therefore the sequence Concat$(\overline{p_m(n)} : n \in \mathbb{N})$ is D-normal, thus completing the proof of Theorem 3.2.

To see how (6.6) follows from (6.5), observe that, in light of the fact that, for fixed $k \in \mathbb{N}$,

$$\sum_{\gamma\in A_D^k} F_\gamma(\rho_x) = \lambda(\rho_x) - k + 1 = \lambda(\rho_x) + O(1), \tag{6.7}$$

we have, as $x \to \infty$,

$$\begin{aligned}
F_\beta(\rho_x) - \frac{\lambda(\rho_x)}{D^k} &= \frac{F_\beta(\rho_x)D^k - \lambda(\rho_x)}{D^k} \\
&= \frac{F_\beta(\rho_x)D^k - \sum_{\gamma\in A_D^k} F_\gamma(\rho_x) + O(1)}{D^k} \\
&= \frac{1}{D^k}\sum_{\gamma\in A_D^k} (F_\beta(\rho_x) - F_\gamma(\rho_x)) + O(1) \\
&= \frac{1}{D^k} D^k o(\lambda(\rho_x)) = o(\lambda(\rho_x)),
\end{aligned}$$

thus proving (6.6).

Hence, we only need to prove (6.5).

Now, from (6.2), it follows that

$$\varDelta(\rho_x) = \sum_{n \in L_x} \varDelta(\overline{p_m(n)}) + O(x). \tag{6.8}$$

Let us further introduce the sets

$$L_x^{(0)} = \{n \in L_x : qp_m(n) \mid n \text{ for some prime } q \in (p_m(n)/3, 3p_m(n))\},$$
$$L_x^{(1)} = \{n \in L_x : \log p_m(n) \le \sqrt{\log x}\exp(-2\sqrt{x_2}\, g(x))\}.$$

With this notation, in light of Lemma 4.2 and (6.4), we then have

$$\begin{aligned} \sum_{n \in L_x^{(0)} \cup L_x^{(1)}} \log p_m(n) &\le cx \log\log x + x\sqrt{\log x}\, \exp(-2\sqrt{x_2}\, g(x)) \\ &= o(x\sqrt{\log x}\, \exp(-\sqrt{x_2}\, g(x))) = o(\lambda(\rho_x)). \end{aligned} \tag{6.9}$$

Hence, setting $L_x^{(2)} = L_x \setminus (L_x^{(0)} \cup L_x^{(1)})$, it follows from (6.8) and (6.9) that

$$\varDelta(\rho_x) = \sum_{n \in L_x^{(2)}} \varDelta(\overline{p_m(n)}) + o(\lambda(\rho_x)). \tag{6.10}$$

Let us now write each integer $n \in L_x^{(2)}$ as $n = ap_m(n)b$, where

$$P(a) \le p_m(n) \le p(b).$$

Thus setting $M = ab$ and given an arbitrarily small $\varepsilon > 0$, from Theorem 1 we have

$$M \le 2x/e^{(\log x)^{1/2-\varepsilon}}. \tag{6.11}$$

Now, let us fix $M = ab$. It is clear that we may ignore those integers $n \le x$ for which $p_m(n)^2 \mid n$ since there are at most $o(x)$ of them anyway. Once this is done, it is clear that in the factorization $n = ap_m(n)b$, we have $P(a) < p(b)$, so that M determines a and b uniquely. Then, in light of (6.11), we may consider the set

$$\mathcal{E}_M := \{n \in L_x^{(2)} : n = ap_m(n)b = Mp_m(n)\}.$$

Let $n_1 < \cdots < n_H$ be the list of all elements of $\mathcal{E}_M$, and further set $\pi_j = p_m(n_j)$ for $j = 1, \ldots, H$. By construction, it is clear that $\pi_1 < \cdots < \pi_H$, all consecutive primes, and since x/M is large by (6.11), it follows that $\pi_H > (3/2)\pi_1$.

Next, let $\mathcal{K}$ be the set of those M's such that the corresponding set $\mathcal{E}_M$ contains at least one $n \in L_x^{(2)}$, since the others need not be accounted for. Hence, for $ab = M$, we deduce that $\mathcal{E}_M$ contains at least $\pi_1/(2\log \pi_1)$ elements, thus implying that $H \ge \pi_1/(2\log \pi_1)$, provided x is chosen to be large enough.

Using Lemma 4.1, it follows that, when $M \in \mathcal{K}$, we have

$$\frac{1}{H} \sum_{j=1}^{H} \frac{|\Delta(\overline{p_m(n_j)})|}{\log p_m(n_j)} \to 0 \qquad \text{as } x \to \infty.$$

From this, it follows that, for $M \in \mathcal{K}$, there exists a function $\varepsilon_x \to 0$ as $x \to \infty$ such that

$$\sum_{M \in \mathcal{K}} \sum_{n \in \mathcal{E}_M} |\Delta(\overline{p_m(n)})| < \varepsilon_x \sum_{M \in \mathcal{K}} \sum_{n \in \mathcal{E}_M} \lambda(\overline{p_m(n)}). \tag{6.12}$$

Using (6.12), estimate (6.5) follows, thus completing the proof of Theorem 3.2.

7. Proof of Theorem 3.3. We first write

$$E_Q(2x) - E_Q(x) = \sum_{x \le n \le 2x} \log p_m(n) \cdot e(Q(p_m(n))). \tag{7.1}$$

Using the notation introduced in the proof of Theorem 3.2, in the above sum we can drop all $n \in L_x^{(0)} \cup L_x^{(1)}$. It follows that we only need to consider $M \in \mathcal{K}$. Now, for a fixed $M \in \mathcal{K}$, we only need to examine the sum

$$\sum_{j=1}^{H} \log \pi_j \cdot e(Q(\pi_j)),$$

where $\pi_1, \ldots, \pi_H$ are consecutive primes and $\pi_H > (3/2)\pi_1$. Using Lemma 3, we then obtain

$$\Big| \sum_{j=1}^{H} \log \pi_j \cdot e(Q(\pi_j)) \Big| \le \varepsilon_x \Big| \sum_{j=1}^{H} \log \pi_j \Big|.$$

Using this in (7.1), it follows that, as $x \to \infty$,

$$\begin{aligned} |E_Q(2x) - E_Q(x)| &= \Big| \sum_{\substack{x \le n \le 2x \\ n \in L_x^{(2)}}} \log p_m(n) \cdot e(Q(p_m(n))) \Big| + o(T(x)) \\ &\le \varepsilon_x T(x) + o(T(x)) = o(T(x)), \end{aligned}$$

as requested.

8. Final remarks. Instead of considering the middle prime factor of an integer, that is the prime factor whose rank amongst the $\omega(n)$ distinct prime factors of an integer n is the $\lfloor \frac{1}{2}\omega(n) \rfloor$th one, we could have also studied the prime factor whose rank is the $\lfloor \alpha\omega(n) \rfloor$th one, for any given real number $\alpha \in (0,1)$. In this more general case, say with $p^{(\alpha)}(n)$ in place of $p_m(n)$, the same type of results as above would also hold, meaning in particular that $\log p^{(\alpha)}(n)$ would be close to $\log^\alpha n$ instead of $\sqrt{\log n}$.

Acknowledgments. The research of the first author was partly supported by a grant from NSREC. The research of the second author was supported by the Hungarian and Vietnamese TET 10-1-2011-0645.

REFERENCES

[1] N. L. Bassily and I. Kátai, *Distribution of consecutive digits in the q-ary expansions of some sequences of integers*, J. Math. Sci. 78 (1996), 11–17.

[2] J. M. De Koninck and I. Kátai, *On a problem on normal numbers raised by Igor Shparlinski*, Bull. Austral. Math. Soc. 84 (2011), 337–349.

[3] J. M. De Koninck and I. Kátai, *Some new methods for constructing normal numbers*, Ann. Sci. Math. Québec 36 (2012), 349–359.

[4] J. M. De Koninck and I. Kátai, *Construction of normal numbers using the distribution of the kth largest prime factor*, Bull. Austral. Math. Soc. 88 (2013), 158–168.

[5] J. M. De Koninck and I. Kátai, *Using large prime divisors to construct normal numbers*, Ann. Univ. Sci. Budapest. Sect. Comput. 39 (2013), 45–62.

[6] J. M. De Koninck and I. Kátai, *Normal numbers generated using the smallest prime factor function*, preprint.

[7] J. M. De Koninck and F. Luca, *On the middle prime factor of an integer*, J. Integer Seq. 13 (2013), art. 13.5.5.

[8] I. M. Vinogradov, *The Method of Trigonometrical Sums in the Theory of Numbers*, Dover Publ., Mineola, NY, 2004.

Jean-Marie De Koninck
Département de mathématiques et de statistique
Université Laval
Québec, QC G1V 0A6, Canada
E-mail: jmdk@mat.ulaval.ca

Imre Kátai
Computer Algebra Department
Eötvös Lorand University
Pázmány Péter sétány 1/C
1117 Budapest, Hungary
E-mail: katai@compalg.inf.elte.hu

Received 7 December 2013;
revised 22 January 2014 (6094)