Jean-Marie De Koninck and Imre Kátai

The number of large prime factors of integers and normal numbers

2015, p. 5-12.

<http://pmb.cedram.org/item?id=PMB_2015____5_0>
THE NUMBER OF LARGE PRIME FACTORS OF INTEGERS AND NORMAL NUMBERS

by

Jean-Marie De Koninck and Imre Kátai

Abstract. — In a series of papers, we constructed large families of normal numbers using the concatenation of the values of the largest prime factor $P(n)$, as $n$ runs through particular sequences of positive integers. A similar approach using the smallest prime factor function also allowed for the construction of normal numbers. Letting $\omega(n)$ stand for the number of distinct prime factors of the positive integer $n$, we then showed that the concatenation of the successive values of $|\omega(n) - \lfloor \log \log n \rfloor|$ in a fixed base $q \geq 2$, as $n$ runs through the integers $n \geq 3$, yields a normal number. Here we prove the following. Let $q \geq 2$ be a fixed integer. Given an integer $n \geq n_0 = \max(q, 3)$, let $N$ be the unique positive integer satisfying $q^N \leq n < q^{N+1}$ and let $h(n, q)$ stand for the residue modulo $q$ of the number of distinct prime factors of $n$ located in the interval $[\log N, N]$. Setting $x_N := e^N$, we then create a normal number in base $q$ using the concatenation of the numbers $h(n, q)$, as $n$ runs through the integers $\geq x_{n_0}$.

Résumé. — Dans une série d’articles, nous avons construit de grandes familles de nombres normaux en utilisant la concaténation des valeurs successives du plus grand facteur premier $P(n)$, où $n$ parcourt certaines suites d’entiers positifs. Une approche similaire en utilisant la fonction plus petit facteur premier nous a aussi permis de construire d’autres familles de nombres normaux. En désignant par $\omega(n)$ le nombre de nombres premiers distincts de $n$, nous avons montré que la concaténation des valeurs successives de $|\omega(n) - \lfloor \log \log n \rfloor|$ dans une base fixe $q \geq 2$, où $n$ parcourt les entiers $n \geq 3$, donne place à un nombre normal. Ici, nous démontrons le résultat suivant. Soit $q \geq 2$ un entier fixe. Étant donné un entier $n \geq n_0 = \max(q, 3)$, soit $N$ l’unique entier positif satisfaisant $q^N \leq n < q^{N+1}$ et désignons par $h(n, q)$ le résidu modulo $q$ du nombre de facteurs premiers distincts de $n$ situés dans l’intervalle $[\log N, N]$. En posant $x_N := e^N$, nous créons alors un nombre normal dans la base $q$ en utilisant la concaténation des nombres $h(n, q)$, où $n$ parcourt les entiers $\geq x_{n_0}$.


Key words and phrases. — Normal numbers, number of prime factors.

Acknowledgements. — The research of the first author was supported in part by a grant from NSERC. The research of the second author was supported by the Hungarian and Vietnamese TET 10-1-2011-0645.
1. Introduction

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

Even though constructing specific normal numbers is a very difficult problem, several authors picked up this challenge. One of the first was Champernowne [2] who, in 1933, showed that the number made up of the concatenation of the natural numbers, namely the number

\[
0.123456789101112131415161718192021\ldots,
\]

is normal in base 10. In 1946, Copeland and Erdős [4] proved that the same is true if one replaces the sequence of natural numbers by the sequence of primes, namely for the number

\[
0.23571113171923293137\ldots
\]

In the same paper, they conjectured that if \( f(x) \) is any nonconstant 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. Six years later, Davenport and Erdős [5] proved this conjecture.

Since then, many others have constructed various families of normal numbers. To name only a few, let us mention Nakai and Shiokawa [15], Madritsch, Thuswaldner and Tichy [14] and finally Vandehey [17]. More examples of normal numbers as well as numerous references can be found in the recent book of Bugeaud [1].

In a series of papers, we also constructed large families of normal numbers using the distribution of the values of \( P(n) \), the largest prime factor function (see [6], [7], [8] and [9]). Recently in [10], we showed how the concatenation of the successive values of the smallest prime factor \( p(n) \), as \( n \) runs through the positive integers, can also yield a normal number. Letting \( \omega(n) \) stand for the number of distinct prime factors of the positive integer \( n \), we then showed that the concatenation of the successive values of \([\omega(n)] - \lfloor \log \log n \rfloor \) in a fixed base \( q \geq 2 \), as \( n \) runs through the integers \( n \geq 3 \), yields a normal number.

Given an integer \( N \geq 1 \), for each integer \( n \in J_N := (e^N, e^{N+1}) \), let \( q_N(n) \) be the smallest prime factor of \( n \) which is larger than \( N \); if no such prime factor exists, set \( q_N(n) = 1 \). Fix an integer \( Q \geq 3 \) and consider the function \( f(n) = f_Q(n) \) defined by \( f(n) = \ell \) if \( n \equiv \ell \mod(Q) \) with \( (\ell, Q) = 1 \) and by \( f(n) = \epsilon \) otherwise, where \( \epsilon \) stands for the empty word. Then consider the sequence \( (\kappa(n))_{n \geq 3} = (\kappa_Q(n))_{n \geq 3} \) defined by \( \kappa(n) = f(q_N(n)) \) if \( n \in J_N \) with \( q_N(n) \) is positive but \( \kappa(n) = \epsilon \) if \( n \in J_N \) with \( q_N(n) = 1 \). Then, given an integer \( N \geq 1 \) and writing \( J_N = \{j_1, j_2, j_3, \ldots \} \), consider the concatenation of the numbers \( \kappa(j_1), \kappa(j_2), \kappa(j_3), \ldots \), that is define

\[
\theta_N := \text{Concat}(\kappa(n) : n \in J_N) = 0.\kappa(j_1)\kappa(j_2)\kappa(j_3)\ldots.
\]

Then, set \( \alpha_Q := \text{Concat}(\theta_N : N = 1, 2, 3, \ldots) \) and let \( B_Q = \{\ell_1, \ell_2, \ldots, \ell_{\varphi(Q)}\} \) be the set of reduced residues modulo \( Q \), where \( \varphi \) stands for the Euler function. In [11], we showed that \( \alpha_Q \) is a normal sequence over \( B_Q \), that is, the real number \( 0.\alpha_Q \) is a normal number over \( B_Q \).

Here we prove the following. Let \( q \geq 2 \) be a fixed integer. Given an integer \( n \geq n_0 = \max(q, 3) \), let \( N \) be the unique positive integer satisfying \( q^N \leq n < q^{N+1} \) and let \( h(n, q) \) stand for the residue modulo \( q \) of the number of distinct prime factors of \( n \) located in the interval \([\log N, N] \).

Setting \( x_N := e^N \), we then create a normal number in base \( q \) using the concatenation of the numbers \( h(n, q) \), as \( n \) runs through the integers \( n \geq x_{n_0} \).
2. The main result

**Theorem 2.1.** — Let \( q \geq 2 \) be a fixed integer. Given an integer \( n \geq n_0 = \max(q, 3) \), let \( N \) be the unique positive integer satisfying \( q^N \leq n < q^{N+1} \) and let \( h(n, q) \) stand for the residue modulo \( q \) of the number of distinct prime factors of \( n \) located in the interval \([\log N, N]\). For each integer \( N \geq 1 \), set \( x_N := e^N \). Then, Concat\((h(n, q) : x_{n_0} \leq n \in \mathbb{N})\) is a \( q \)-ary normal sequence.

**Proof.** — For each integer \( N \geq 1 \), let \( J_N = (x_N, x_{N+1}) \). Further let \( S_N \) stand for the set of primes located in the interval \([\log N, N]\) and \( T_N \) for the product of the primes in \( S_N \). Let \( n_0 = \max(q, 3) \). Given a large integer \( N \), consider the function
\[
(1) \quad f(n) = f_N(n) = \sum_{p|n \log N \leq p \leq N} 1.
\]
Let us further introduce the following sequences:
\[
U_N = \text{Concat} \left( h(n, q) : n \in J_N \right),
\]
\[
V_\infty = \text{Concat} \left( U_N : N \geq n_0 \right) = \text{Concat} \left( h(n, q) : n \geq x_{n_0} \right),
\]
\[
V_x = \text{Concat} \left( h(n, q) : x_{n_0} \leq n \leq x \right).
\]
Let us set \( A_q := \{0, 1, \ldots, q-1\} \). If we fix an arbitrary integer \( r \), it is sufficient to prove that given any particular word \( w \in A_q^r \), the number of occurrences \( F_w(V_x) \) of \( w \) in \( V_x \) satisfies
\[
(2) \quad F_w(V_x) = (1 + o(1)) \frac{x}{q^r} \quad (x \to \infty).
\]
For each integer \( r \geq 1 \), considering the polynomial
\[
Q_r(u) = u(u+1) \cdots (u+r-1).
\]
and letting
\[
\rho_r(d) = \#\{u \mod d : Q_r(u) \equiv 0 \mod d\},
\]
it is clear that, since \( N \) is large,
\[
(3) \quad \rho_r(p) = r \quad \text{if} \ p \in S_N.
\]
Observe that it follows from the Turán-Kubilius Inequality (see for instance Theorem 7.1 in the book of De Koninck and Luca [12]), that for some positive constant \( C \),
\[
(4) \quad \sum_{n \in J_N} (f(n) - \log \log N)^2 \leq C e^N \log \log N.
\]
Letting \( \varepsilon_N = 1/\log \log N \), it follows from (4) that
\[
(5) \quad \frac{1}{x_N} \#\{n \in J_N : |f(n) - \log \log N| > \frac{1}{\varepsilon_N} \sqrt{\log \log N} \} \to 0 \quad (N \to \infty).
\]
This means that in the estimation of \( F_w(V_x) \), we may ignore those integers \( n \) appearing in the concatenation \( h(2, q)h(3, q) \ldots h([x], q) \) for which the corresponding \( f(n) \) is “far” from \( \log \log N \) in the sense described in (5).

Let \( X \) be a large number. Then there exists a large integer \( N \) such that \( \frac{X}{e} < x_N \leq X \). Letting
\[
L = \left\lceil \frac{X}{e}, X \right\rceil,
\]
we write...
say, and $\lambda(L_i)$ for the length of the interval $L_i$ for $i = 1, 2$.

Given an arbitrary function $\delta_N$ which tends to 0 arbitrarily slowly, it is sufficient to consider those $L_1$ and $L_2$ such that

$$\lambda(L_1) \geq \delta_N X \quad \text{and} \quad \lambda(L_2) \geq \delta_N X.$$  

The reason for this is that those $n \in L_1$ (resp. $n \in L_2$) for which $\lambda(L_1) < \delta_N X$ (resp. $\lambda(L_2) < \delta_N X$) are $o(x)$ in number and can therefore be ignored in the proof of (2).

Let us first consider the set $L_2$. We start by observing that any subword taken in the concatenation $h(n, q)h(n + 1, q)\ldots h(n + r - 1, q)$ is made of co-prime divisors of $T_N$ (since no two members of the sequence $h(n, q), h(n + 1, q), \ldots, h(n + r - 1, q)$ of $r$ elements may have a common prime divisor $p > \log N$). So, let $d_0, d_1, \ldots, d_{r-1}$ be co-prime divisors of $T_N$ and let $B_N(L_2; d_0, d_1, \ldots, d_{r-1})$ stand for the number of those $n \in L_2$ for which $d_j \mid n + j$ for $j = 0, 1, \ldots, r - 1$ and such that $\left( \frac{Q_r(n)}{d_0d_1\ldots d_{r-1}} \right) = 1$. We can assume that each of the $d_j$'s is squarefree, since the number of those $n + j \leq X$ for which $p^2 \mid n + j$ for some $p > \log N$ is $\ll X \sum_{p > \log N} \frac{1}{p^2} = o(X)$.

In light of (4), we may assume that

$$\omega(d_j) \leq 2\log \log N \quad \text{for} \quad j = 0, 1, \ldots, r - 1.$$  

By using the Eratosthenian sieve (see for instance the book of De Koninck and Luca [12]) and recalling that condition (6) ensures that $X - x_N$ is large, we obtain that, as $N \rightarrow \infty$,

$$B_N(L_2; d_0, d_1, \ldots, d_{r-1}) = \frac{X - x_N}{d_0d_1\ldots d_{r-1}} \prod_{p \mid T_N/(d_0d_1\ldots d_{r-1})} \left( 1 - \frac{r}{p} \right)$$  

$$+ o \left( \frac{x_N}{d_0d_1\ldots d_{r-1}} \prod_{p \mid T_N/(d_0d_1\ldots d_{r-1})} \left( 1 - \frac{r}{p} \right) \right).$$  

Letting $\theta_N := \prod_{p \mid T_N} \left( 1 - \frac{r}{p} \right)$, one can easily see that

$$\theta_N = (1 + o(1)) \left( \frac{\log \log N}{\log N} \right)^r \quad (N \rightarrow \infty).$$  

Let us also introduce the strongly multiplicative function $\kappa$ defined on primes $p$ by $\kappa(p) = p - r$. Then, (8) can be written as

$$B_N(L_2; d_0, d_1, \ldots, d_{r-1}) = \frac{X - x_N}{\kappa(d_0)\kappa(d_1)\ldots\kappa(d_{r-1})} \theta_N + o \left( \frac{x_N}{\kappa(d_0)\kappa(d_1)\ldots\kappa(d_{r-1})} \theta_N \right)$$  

as $N \rightarrow \infty$. For each integer $N > e^\varepsilon$, let

$$R_N := \left[ \log \log N - \frac{\sqrt{\log \log N}}{\varepsilon}, \log \log N + \frac{\sqrt{\log \log N}}{\varepsilon} \right].$$
Let \( \ell_0, \ell_1, \ldots, \ell_{r-1} \) be an arbitrary collection of non negative integers \( < q \). Note that there are \( q^r \) such collections. Our goal is to count how many times, amongst the integers \( n \in \mathcal{L}_2 \), we have \( f(n+j) \equiv \ell_j \pmod{q} \) for \( j = 0, 1, \ldots, r-1 \). In light of (5), we only need to consider those \( n \in \mathcal{L}_2 \) for which

\[
f(n+j) \in R_N \quad (j = 0, 1, \ldots, r-1).
\]

Let

\[
\mathcal{S}(\ell_0, \ell_1, \ldots, \ell_{r-1}) := \sum_{\substack{d_j \equiv \ell_j \pmod{q} \\
\text{j=1}^{1} \ldots \text{r-1}\n d_j \in \mathbb{T}_N}} 1, 
\]

where the star over the sum indicates that the summation runs only on those \( d_j \) satisfying \( f(d_j) \in R_N \) for \( j = 0, 1, \ldots, r-1 \).

From (10), we therefore obtain that

\[
\#\{n \in \mathcal{L}_2 : f(n+j) \equiv \ell_j \pmod{q}, j = 0, 1, \ldots, r-1\} = (X - x_N)^{\theta_N} \mathcal{S}(\ell_0, \ell_1, \ldots, \ell_{r-1}) + o(x_N^{\theta_N} \mathcal{S}(\ell_0, \ell_1, \ldots, \ell_{r-1})) \\
\text{(12)}
\]
as \( N \to \infty \). Let us now introduce the function

\[
\eta = \eta_N = \sum_{p \mid T_N} \frac{1}{\kappa(p)}.
\]

Observe that, as \( N \to \infty \),

\[
\eta = \sum_{\log N \leq p \leq N} \frac{1}{p(1 - r/p)} = \sum_{\log N \leq p \leq N} \frac{1}{p} + O \left( \sum_{\log N \leq p \leq N} \frac{1}{p^2} \right) \\
\text{(13)} = \log \log N - \log \log \log N + o(1) + O \left( \frac{1}{\log N} \right).
\]

From the definition (11), one easily sees that

\[
\mathcal{S}(\ell_0, \ell_1, \ldots, \ell_{r-1}) = (1 + o(1)) \sum_{\substack{t_j \equiv \ell_j \pmod{q} \\
t_j \in R_N}} \frac{\eta_j^{t_0+t_1+\cdots+t_{r-1}}}{t_0!t_1!\cdots t_{r-1}!} \\
\text{(14)} \quad (N \to \infty),
\]

where we ignore in the denominator of the summands the factors \( \kappa(p)^a \) (with \( a \geq 2 \)) since their contribution is negligible.

Moreover, for \( t \in R_N \), one can easily establish that

\[
\frac{\eta^{t+1}}{(t+1)!} = (1 + o(1)) \frac{\eta^t}{t!} \\
\text{(15)} \quad (N \to \infty)
\]

and consequently that, for each \( j \in \{0, 1, \ldots, r-1\} \),

\[
\sum_{\substack{t_j \equiv \ell_j \pmod{q} \\
t_j \in R_N}} \frac{\eta_j^{t}}{t!} = (1 + o(1)) \frac{1}{q} \sum_{t \in R_N} \frac{\eta^t}{t!} = (1 + o(1)) \frac{e^\eta}{q} \\
\text{(15)} \quad (N \to \infty).
\]
Using (15) in (14), we obtain that
\begin{equation}
S(\ell_0, \ell_1, \ldots, \ell_{r-1}) = (1 + o(1)) \frac{e^{\eta r}}{q^r} \quad (N \to \infty) .
\end{equation}
Combining (12) and (16), we obtain that
\begin{equation}
\# \{ n \in \mathcal{L}_2 : f(n+j) \equiv \ell_j \mod q, \ j = 0, 1, \ldots, r-1 \} \\
= (X - x_N) \theta_N \frac{e^{\eta r}}{q^r} + o \left( x_N \theta_N \frac{e^{\eta r}}{q^r} \right) \\
= \frac{X - x_N}{q^r} + o \left( \frac{x_N}{q^r} \right) \quad (N \to \infty),
\end{equation}
where we used (9) and (13).
Since the first term on the right hand side of (17) does not depend on the particular collection \( \ell_0, \ell_1, \ldots, \ell_{r-1} \), we may conclude that the frequency of those integers \( n \in \mathcal{L}_2 \) for which \( f(n+j) \equiv \ell_j \mod q \) for \( j = 0, 1, \ldots, r-1 \) is the same independently of the choice of \( \ell_0, \ell_1, \ldots, \ell_{r-1} \).
The case of those \( n \in \mathcal{L}_1 \) can be handled in a similar way.
We have thus shown that the number of occurrences of any word \( w \in A_q \) in \( h(n,q)h(n+1,q) \ldots h(n+r-1,q) \) as \( n \) runs over the \( \lfloor X / e \rfloor \) elements of \( \mathcal{L} \) is \( (1 + o(1)) \frac{(X - X/e)}{q^r} \).
Repeating this for each of the intervals
\[ \left[ \frac{X}{e^j+1}, \frac{X}{e^j} \right) \quad (j = 0, 1, \ldots, \lfloor \log x \rfloor), \]
we obtain that the number of occurrences of \( w \) for \( n \leq x \) is \( (1 + o(1)) \frac{x}{q^r} \), as claimed.
The proof of (2) is thus complete and the Theorem is proved.

\[ \square \]

3. Final remarks

First of all, let us first mention that our main result can most likely be generalized in order that the following statement will be true:

Let \( a(n) \) and \( b(n) \) be two monotonically increasing sequences of \( n \) for \( n = 1, 2, \ldots \) such that \( n/b(n) \), \( b(n)/a(n) \) and \( a(n) \) all tend to infinity monotonically as \( n \to \infty \). Let \( f(n) \) stand for the number of prime divisors of \( n \) located in the interval \( [a(n), b(n)] \) and let \( h(n,q) \) be the residue of \( f(n) \) modulo \( q \); then, the sequence \( h(n,q), n = 1, 2, \ldots, \) is a \( q \)-ary normal sequence.

Secondly, let us first recall that it was proven by Pillai [16] (with a more general result by Delange [13]) that the values of \( \omega(n) \) are equally distributed over the residue classes modulo \( q \) for every integer \( q \geq 2 \), and that the same holds for the function \( \Omega(n) \), where \( \Omega(n) := \sum_{p^\alpha || n} \alpha \).
We believe that each of the sequences Concat(\( \omega(n) \) mod \( q \) : \( n \in \mathbb{N} \)) and Concat(\( \Omega(n) \) (mod \( q \) : \( n \in \mathbb{N} \)) represents a normal sequence for each base \( q = 2, 3, \ldots \). However, the proof
of these statements could be very difficult to obtain. Indeed, in the particular case \( q = 2 \), such a result would imply the famous Chowla conjecture

\[
\lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} \lambda(n)\lambda(n + a_1) \ldots \lambda(n + a_k) = 0,
\]

where \( \lambda(n) := (-1)^{\Omega(n)} \) is known as the Liouville function and where \( a_1, a_2, \ldots, a_k \) are \( k \) distinct positive integers (see Chowla [3]).

Thirdly, we had previously conjectured that, given any integer \( q \geq 2 \) and letting \( \text{res}_q(n) \) stand for the residue of \( n \) modulo \( q \), it may not be possible to create an infinite sequence of positive integers \( n_1 < n_2 < \cdots \) such that

\[
0.\text{Concat}(\text{res}_q(n_j) : j = 1, 2, \ldots)
\]

is a \( q \)-normal number. However, we now have succeeded in creating such a monotonic sequence. It goes as follows. Let us define the sequence \( (m_k)_{k \geq 1} \) by

\[
m_k = f(k) + k!,
\]

where \( f \) is the function defined in (1). In this case, we obtain that

\[
m_{k+1} - m_k = k! \cdot k + f(k + 1) - f(k),
\]

a quantity which is positive for all integers \( k \geq 1 \) provided

(18) \quad \quad f(k + 1) - f(k) > -k! \cdot k,

that is if

(19) \quad \quad f(k) < k! \cdot k.

But since we trivially have

\[
f(k) \leq \omega(k) \leq 2 \log k \leq k! \cdot k,
\]

then (19) follows and therefore (18) as well.

Hence, in light of Theorem 2.1, if we choose \( n_k = m_k \), our conjecture is disproved.

4. Acknowledgment

The authors are grateful to Professor Andrzej Schinzel, since the problem examined here originated from an email exchange between the second author and Professor Schinzel. The authors are also grateful to the referee whose comments helped improve the quality of this paper.

References


April 3, 2014

JEAN-MARIE DE KONINCK, Département de mathématiques et de statistique, Université Laval,
1045 Avenue de la médecine, Québec, Québec G1V 0A6, Canada • E-mail : jmdk@mat.ulaval.ca
Url : www.jeanmariedekoninck.mat.ulaval.ca

IMRE KÁTAI, Computer Algebra Department, Eötvös Loránd University, 1117 Budapest, Pézmány Péter Sétány 1/C, Hungary • E-mail : katai@compalg.inf.elte.hu

Publications mathématiques de Besançon – 2015