Prime-like sequences leading to the construction of normal numbers

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

Abstract

Given an integer $q \geq 2$, a *q-normal number* is an irrational number η such that any preassigned sequence of k digits occurs in the q-ary expansion of η at the expected frequency, namely $1/q^k$. Given an integer $q \geq 3$, we consider the sequence of primes reduced modulo q and examine various possibilities of constructing normal numbers using this sequence. We create a sequence of independent random variables that mimics the sequence of primes and then show that for almost all outcomes this allows to obtain a normal number.

AMS Subject Classification numbers: 11K16, 11A41, 11N13
Key words: normal numbers, primes, arithmetic progression

Édition du 11 septembre 2012

1 Introduction

Given an integer $q \geq 2$, a *q-normal number*, or simply a *normal number*, is an irrational number whose q-ary expansion is such that any preassigned sequence, of length $k \geq 1$, of base q digits from this expansion, occurs at the expected frequency, namely $1/q^k$.

In earlier papers [3], [4], [5], [6], we used the complexity of the factorization of integers to create large families of normal numbers. In this paper, given an integer $q \geq 3$, we consider the sequence of primes reduced modulo q and examine various possibilities of constructing normal numbers using this sequence.

Let $A_q := \{0, 1, \ldots, q-1\}$. Given an integer $t \geq 1$, an expression of the form $i_1 i_2 \ldots i_t$, where each $i_j \in A_q$, is called a *finite word* of length t. The symbol Λ will denote the *empty word*. We let A_q^t stand for the set of all words of length t and A_q^* stand for the set of all the words regardless of their length. An infinite sequence of digits $a_1 a_2 \ldots$, where each $a_i \in A_q$, is called an *infinite word*.

An infinite sequence of base q digits $a_1 a_2 \ldots$ will be said to be a *normal sequence* if any preassigned sequence of k digits occurs at the expected frequency of $1/q^k$.

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

Given a fixed integer $q \geq 3$, let

$$f_q(n) = \begin{cases} \Lambda & \text{if } (n,q) \neq 1, \\ \ell & \text{if } n \equiv \ell \pmod{q}, \quad (\ell,q) = 1. \end{cases}$$

Further, letting φ stand for the Euler function, set

$$B_{\varphi(q)} = \{\ell_1, \ldots, \ell_{\varphi(q)}\}$$

be the set of reduced residues modulo q.

Let $\wp$ stand for the set of all primes, writing $p_1 < p_2 < \cdots$ for the sequence of consecutive primes, and consider the infinite word

$$\xi_q = f_q(p_1) f_q(p_2) f_q(p_3) \ldots$$

We first state the following conjecture.

Conjecture 1. *The word ξ_q is a normal sequence over $B_{\varphi(q)}$ in the sense that given any integer $k \geq 1$ and any word $\beta = r_1 \ldots r_k \in B^k_{\varphi(q)}$, then, setting*

$$\xi_q^{(N)} = f_q(p_1) f_q(p_2) \ldots f_q(p_N) \quad \textit{for each} \quad N \in \mathbb{N}$$

and

$$M_N(\xi_q|\beta) := \#\{(\gamma_1, \gamma_2) | \xi_q^{(N)} = \gamma_1 \beta \gamma_2\},$$

we have

$$\lim_{N\to\infty} \frac{M_N(\xi_q|\beta)}{N} = \frac{1}{\varphi(q)^k}.$$

Now, with the above notation, consider the following weaker conjecture.

Conjecture 2. *For every finite word β, there exists a positive integer N such that $M_N(\xi_q|\beta) > 0$.*

Remark 1. *Observe that, in 2000, Shiu [10] provided some hope in the direction of a proof of this last conjecture by proving that given any positive integer k, there exists a string of congruent primes of length k, that is a set of consecutive primes $p_{n+1} < p_{n+2} < \cdots < p_{n+k}$ (where p_i stands for the i-th prime) such that*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod{q},$$

for some positive integer n, for any given modulus q and positive integer a relatively prime with q.

Let ε_n be a real function which tends monotonically to 0 as $n \to \infty$ but in such a way that $(\log\log n)\varepsilon_n \to \infty$ as $n \to \infty$. Letting $p(n)$ stand for the smallest prime factor of n, consider the set

$$\mathcal{N}^{(\varepsilon_n)} := \{n \in \mathbb{N} : p(n) > n^{\varepsilon_n}\} = \{n_1, n_2, \ldots\}. \tag{1.1}$$

We then have the following conjecture.

Conjecture 3. *Let $n_1 < n_2 < \cdots$ be the sequence defined in (1.1). Then the infinite word*

$$\xi_q := f_q(n_1) f_q(n_2) \ldots$$

is a normal sequence over the set $\{\ell \mod q : (\ell, q) = 1\}$.

Although the problem of generating normal numbers using the sequence of primes does seem inaccessible, we will nevertheless manage to create large families of normal numbers, in the direction of Conjectures 1, 2 and 3, but this time using prime-like sequences.

2 Main results

Theorem 1. *Let $n_1 < n_2 < \cdots$ be the sequence defined in (1.1). Then the infinite word*

$$\eta_q := res_q(n_1) res_q(n_2) \ldots,$$

where $res_q(n) = \ell$ if $n \equiv \ell \pmod{q}$, contains every finite word whose digits belong to $B_{\varphi(q)}$ infinitely often.

Remark 2. *It is now convenient to recall a famous conjecture concerning the distribution of primes.*

Let $F_1, \ldots, F_g$ be distinct irreducible polynomials in $\mathbb{Z}[x]$ (with positive leading coefficients) and assume that the product $F := F_1 \cdots F_g$ has no fixed prime divisor. Then the famous Hypothesis H of Schinzel and Sierpinski [9] states that there exist infinitely many integers n such that each $F_i(n)$ ($i = 1, \ldots, g$) is a prime number. The following quantitative form of Hypothesis H was later given by Bateman and Horn ([1],[2]):

> *(*Bateman-Horn Hypothesis*) If $Q(F_1, \ldots, F_g; x)$ stands for the number of positive integers $n \le x$ such that each $F_i(n)$ ($i = 1, \ldots, g$) is a prime number, then*
>
> $$Q(F_1, \ldots, F_g; x) = (1 + o(1)) \frac{C(F_1, \ldots, F_g)}{h_1 \cdots h_g} \frac{x}{\log^g x} \qquad (x \to \infty),$$
>
> *where $h_i = deg F_i$ and*
>
> $$C(F_1, \ldots, F_g) = \prod_p \left(\left(1 - \frac{1}{p}\right)^{-g} \left(1 - \frac{\rho(p)}{p}\right) \right),$$

with $\rho(p)$ denoting the number of solutions of $F_1(n)\cdots F_g(n) \equiv 0 \pmod{p}$.

Theorem 2. *Let β be an arbitrary word belonging to $B^k_{\varphi(q)}$ and let ξ_q be defined as in Conjecture 3. If the Bateman-Horn Hypothesis holds, then*

$$M_N(\xi_q|\beta) \to \infty \qquad \text{as } N \to \infty.$$

Let

$$\lambda_m = \begin{cases} 0 & \text{if } m = 1, 2, \ldots, 10, \\ 1/\log m & \text{if } m \geq 11. \end{cases}$$

Let ξ_m be a sequence of independent random variables defined by $P(\xi_m = 1) = \lambda_m$ and $P(\xi_m = 0) = 1 - \lambda_m$. Let Ω be the set of all possible events ω in this probability space.

Let ω be a particular outcome, say $m_1, m_2, \ldots$, that is one for which $\xi_{m_j} = 1$ for $j = 1, 2, \ldots$ and $\xi_\ell = 0$ if $\ell \notin \{m_1, m_2, \ldots\}$. Now, for a fixed integer $q \geq 3$, set $\mathrm{res}_q(m) = \ell$ if $m \equiv \ell \pmod{q}$, with $\ell \in A_q$. Then, let $\eta_q(\omega)$ be the real number whose q-ary expansion is given by

$$\eta_q(\omega) = 0.\mathrm{res}_q(m_1)\mathrm{res}_q(m_2)\ldots$$

We then have the following result.

Theorem 3. *The number $\eta_q(\omega)$ is a q-normal number for almost all outcomes ω.*

3 Preliminary results

For here on, the letter c will be used to denote a positive constant, but not necessarily the same at each occurrence.

Lemma 1. *Let $q \geq 2$, $k \geq 1$ and $M \geq 1$ be fixed integers. Given any nonnegative integer $n < q^M$, write its q-ary expansion as*

$$n = \sum_{j=0}^{M-1} \varepsilon_j(n) q^j, \qquad \varepsilon_j(n) \in A_q$$

and, given any word $\alpha = b_1 \ldots b_k \in A_q^k$, set

$$E_\alpha(n) := \#\{j \in \{0, 1, \ldots, M-k\} : \varepsilon_j(n) \ldots \varepsilon_{j+k-1}(n) = \alpha\}.$$

Then, there exists a constant $c = c(k, q)$ such that

$$\sum_{0 \leq n < q^M} \left(E_\alpha(n) - \frac{M}{q^k} \right)^2 \leq c\, q^M M.$$

Proof. Let

$$f(c_1, \ldots, c_k) = \begin{cases} 1 & \text{if } (c_1, \ldots, c_k) = (b_1, \ldots, b_k), \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\Sigma_1 := \sum_{0 \le n < q^M} E_\alpha(n) = \sum_{0 \le n < q^M} \sum_{j=0}^{M-k-1} f(\varepsilon_j(n), \ldots, \varepsilon_{j+k-1}(n)) = q^{M-k}(M-k).$$

Similarly,

$$\begin{aligned} \Sigma_2 &:= \sum_{0 \le n < q^M} E_\alpha(n)^2 \\ &= \sum_{0 \le n < q^M} \sum_{j_1=0}^{M-k-1} \sum_{j_2=0}^{M-k-1} f(\varepsilon_{j_1}(n), \ldots, \varepsilon_{j_1+k-1}(n)) \cdot f(\varepsilon_{j_2}(n), \ldots, \varepsilon_{j_2+k-1}(n)) \\ &= \sum_{0 \le n < q^M} \sum_{|j_1 - j_2| \le k} f(\varepsilon_{j_1}(n), \ldots, \varepsilon_{j_1+k-1}(n)) \cdot f(\varepsilon_{j_2}(n), \ldots, \varepsilon_{j_2+k-1}(n)) \\ &\quad + \sum_{0 \le n < q^M} \sum_{|j_1 - j_2| > k} f(\varepsilon_{j_1}(n), \ldots, \varepsilon_{j_1+k-1}(n)) \cdot f(\varepsilon_{j_2}(n), \ldots, \varepsilon_{j_2+k-1}(n)) \\ &= \Sigma_{2,1} + \Sigma_{2,2}, \end{aligned}$$

say.

On the one hand, it is clear that

$$0 \le \Sigma_{2,1} \le (2k+1) q^{M-k}(M-k) \le c q^M M. \tag{3.1}$$

On the other hand, to estimate $\Sigma_{2,2}$, first observe that for fixed j_1, j_2 with $|j_1 - j_2| > k$, we have to sum 1 over those $n \in [0, q^M - 1[$ for which

$$\varepsilon_{j_1}(n) \ldots \varepsilon_{j_1+k-1}(n) = \alpha = \varepsilon_{j_2}(n) \ldots \varepsilon_{j_2+k-1}(n).$$

But this occurs exactly for q^{M-2k} many n's. Thus,

$$\Sigma_{2,2} = q^{M-2k} \sum_{\substack{|j_1 - j_2| > k \\ 0 \le j_1, j_2 \le M-k-1}} 1 = q^{M-2k} M^2 + O(q^M M). \tag{3.2}$$

In light of (3.1) and (3.2), it follows that

$$\begin{aligned} \sum_{0 \le n < q^M} \left(E_\alpha(n) - \frac{M}{q^k} \right)^2 &= \Sigma_2 - 2\frac{M}{q^k}\Sigma_1 + \frac{M^2}{q^{2k}} q^M \\ &= q^{M-2k} M^2 + O(q^M M) - 2\frac{M^2}{q^{2k}} q^M + \frac{M^2}{q^{2k}} q^M \\ &= O(q^M M), \end{aligned}$$

thus completing the proof of the lemma. □

Lemma 2. *Given a fixed positive integer R, consider the word $\kappa = c_1 \dots c_R \in A_q^R$. Fix another word $\alpha = b_1 \dots b_k \in A_q^k$, with $k \le R$. Let K_1 stand for the number of solutions (γ_1, γ_2) of $\kappa = \gamma_1 \alpha \gamma_2$, that is the number of those j's for which $c_{j+1} \dots c_{j+k} = \alpha$. Then, given fixed indices $i_1, \dots, i_H$, let K_2 be the number of solutions of $c_{j+1} \dots c_{j+k} = \alpha$ for which $\{j+1, \dots, j+k\} \cap \{i_1, \dots, i_H\} = \emptyset$ holds. Then,*

$$0 \le K_1 - K_2 \le 2kH.$$

Proof. The proof is obvious. □

Lemma 3 (Borel-Cantelli Lemma)**.** *Let $\{E_n\}_{n \in \mathbb{N}}$ be an infinite sequence of events in some probability space. Assuming that the sum of the probabilities of the E_n's is finite, that is, $\sum_{n=1}^{\infty} P(E_n) < +\infty$, then the probability that infinitely many of them occur is 0.*

Proof. For a proof of this result, see the book of Janos Galambos [8]. □

Lemma 4. *Let $F_1, \dots, F_g$ be distinct irreducible polynomials in $\mathbb{Z}[x]$ (with positive leading coefficients) and set $F := F_1 \cdots F_g$. Let $\rho(p)$ stand for the number of solutions of $F(n) \equiv 0 \pmod p$ and assume that $\rho(p) < p$ for all primes p. Write $p(n)$ for the smallest prime factor of the integer $n \ge 2$ and assume that u and x are real numbers satisfying $u \ge 1$ and $x^{1/u} \ge 2$. Then,*

$$\begin{aligned}
&\#\{n \le x : F_i(n) = q_i \text{ for } i = 1, \dots, k\} \\
&= x \prod_{p < x^{1/u}} \left(1 - \frac{\rho(p)}{p}\right) \\
&\quad \times \left\{1 + O_F(\exp(-u(\log u - \log\log 3u - \log k - 2))) + O_F(\exp(-\sqrt{\log x}))\right\}.
\end{aligned}$$

Proof. This is Theorem 2.6 in the book of Halbertsam and Richert [7]. □

4 Proof of Theorem 1

Theorem 1 is essentially a consequence of Lemma 4. Indeed, letting $a_1 < \dots < a_k$ be positive integers coprime to q and considering the product of linear polynomials

$$F(n) := (qn + a_1) \cdots (qn + a_k),$$

we have that

(4.1)

$$\#\{n \in [x, 2x] : p(F(n)) > (2qx + a_k)^{\varepsilon_x}\} = (1 + o(1)) x \prod_{p < x^{\varepsilon_x}} \left(1 - \frac{\rho(p)}{p}\right).$$

If n is counted in (4.1), we certainly have that $p(qn+a_j) > (qn+a_j)^{\varepsilon_{qn+a_j}}$ for $j = 1, \dots, k$. On the other hand, the desired numbers $qn + a_j$, $j = 1, \dots, k$, are consecutive integers with no small prime factors for all but a negligible number. Indeed, if they were not consecutive, then there would be an integer $b \in (a_1, a_k)$ such that $p(qn + b) > x^{\varepsilon_x}$. Set $G_b(n) := qn + b$. Then we would have

(4.2)

$$\#\{n \in [x, 2x] : p(F(n)G_b(n)) > x^{\varepsilon_x}\} = (1 + o(1))x \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho_b(p)}{p}\right),$$

where $\rho_b(p)$ stands for the number of solutions of $F(n)G_b(n) \equiv 0 \pmod p$. Since $\rho(p) = k$ (recall that the F_i's are linear) and $\rho_b(p) = k + 1$ if $p \nmid q$, $p > a_k$, it follows that we have the following two "opposite" inequalities:

$$\begin{aligned} \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho(p)}{p}\right) &\geq C(a_1, \dots, a_k) (\varepsilon_x \log x)^{-k}, \\ \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho_b(p)}{p}\right) &\leq C(a_1, \dots, a_k) (\varepsilon_x \log x)^{-k-1}. \end{aligned}$$

Now, for the choice of b, we clearly have $a_k - a_1 + 1 - k$ possible values. We have thus proved that for every large number x, there is at least one $n \in [x, 2x]$ for which the numbers $qn+a_1, \dots, qn+a_k$ are consecutive integers without small prime factors, that is for which $p(qn + a_j) > (qn + a_j)^{\varepsilon_{qn+a_j}}$, thus completing the proof of Theorem 1.

5 Proof of Theorem 2

The proof of Theorem 2 is almost similar to the one of Theorem 1. Indeed assume that the Bateman-Horn Hypothesis holds (see Remark 2 above). Then, let a_1 be a positive integer such that $a_1 \equiv b_1 \pmod q$ and $a_1 \equiv 0 \pmod D$, where $D = \prod_{\substack{\pi \leq k \\ \pi \nmid q}} \pi$, where π are primes. Similarly, let a_2 be a positive integer such that $a_2 \equiv b_2 \pmod q$ and $a_2 \equiv 0 \pmod D$, with $a_2 > a_1$. Continuing in this manner, that is if $a_1, \dots, a_{\ell-1}$ have been chosen, we let $a_\ell \equiv b_\ell \pmod q$ with $D | a_\ell$ and $a_\ell > a_{\ell-1}$. Then, applying the Bateman-Horn Hypothesis, we get that if $0 < a_1 < \cdots < a_k$ are k integers satisfying $(a_j, q) = 1$ for $j = 1, \dots, k$, then for each positive integer n, setting

$$F(n) = (qn + a_1) \cdots (qn + a_k),$$

letting

$$\rho(m) = \#\{\nu \pmod m : F(\nu) \equiv 0 \pmod m\},$$

so that $\rho(m) = 0$ if $(m, q) > 1$ and $\rho(p) < p$ for each prime p, and further setting

$$\Pi_x := \prod_{\substack{p \in \wp \\ p \leq \sqrt{qx + a_k}}} p,$$

we have that, as $x \to \infty$, letting μ stand for the Moebius function,

$$\begin{aligned}
\sum_{\substack{n \leq x \\ (F(n), \Pi_x) = 1}} 1 &= \sum_{n \leq x} \sum_{\delta | (F(n), \Pi_x)} \mu(\delta) = \sum_{\delta | \Pi_x} \mu(\delta) \sum_{\substack{n \leq x \\ F(n) \equiv 0 \pmod{\delta}}} 1 \\
&= (1 + o(1)) x \sum_{\delta | \Pi_x} \frac{\mu(\delta)\rho(\delta)}{\delta} = (1 + o(1)) x \prod_{p \leq \sqrt{qx + a_k}} \left(1 - \frac{\rho(p)}{p}\right) \\
&= (1 + o(1)) c \frac{x}{\log^k x}, \qquad (5.1)
\end{aligned}$$

where c is a positive constant which depends only on $a_1, \ldots, a_k$.

Now, we can show that almost all prime solutions $\pi_1 < \cdots < \pi_k$ represent a chain of consecutive primes. To see this, assume the contrary, that is that the primes $\pi_1 < \cdots < \pi_k$ are not consecutive, meaning that there exists a prime π satisfying $\pi_1 < \pi < \pi_k$ and $\pi \notin \{\pi_2, \ldots, \pi_{k-1}\}$. Assume that $\pi_\ell < \pi < \pi_{\ell+1}$ for some $\ell \in \{1, \ldots, k-1\}$. We then have

$$\begin{aligned}
\pi_2 &= \pi_1 + a_2 - a_1, \\
\pi_3 &= \pi_1 + a_3 - a_1, \\
\vdots &= \vdots \\
\pi_\ell &= \pi_1 + a_\ell - a_1, \\
\vdots &= \vdots \\
\pi_k &= \pi_1 + a_k - a_1, \\
\pi &= \pi_1 + d, \text{ where } a_\ell - a_1 < d < a_{\ell+1} - a_1.
\end{aligned}$$

We can now find an upper bound for the number of such $k + 1$ tuples. Indeed, by using the Brun-Selberg sieve, one can obtain that the number of such solutions up to x is no larger than $c \dfrac{x}{\log^{k+1} x}$, which in light of (5.1) proves our claim, thus completing the proof of Theorem 2.

6 Proof of Theorem 3

Let $N \geq 3$. Choose a positive integer R which is such that $S := q^N + qR < q^{N+1}$. Then, set

$$Y_{N,S} = \sum_{q^N \leq n < S} \xi_n \quad \text{and} \quad \theta_{N,S} = \sum_{q^N \leq n < S} \xi_n (\xi_{n+1} + \cdots + \xi_{n+q-1}).$$

Then

$$E(Y_{N,S}) = \sum_{q^N \le n < S} \frac{1}{\log n} = \left(\frac{S}{\log S} - \frac{q^N}{\log q^N} \right) + O\left(\frac{q^N}{N^2} \right),$$

while

$$E(Y_{N,S} - E(Y_{N,S}))^2 \le c \sum_{q^N \le n < S} \lambda_n \le c \frac{q^N}{N}.$$

From the Tchebyshev inequality, we then get that

$$P\left(|Y_{N,S} - E(Y_{N,S})| > \sqrt{\frac{q^{N+1}}{\log q^{N+1}}} \right) < \frac{c}{N^2}. \tag{6.1}$$

Similarly

$$P\left(\theta_{N,S} > c_1 \frac{q^N}{N^2} \right) < \frac{1}{N^2}. \tag{6.2}$$

Now, given T integers $n_1 < \cdots < n_T$ located in the interval $[q^N, S-1]$, consider the set

$$\begin{aligned} B_{n_1,\ldots,n_T} = \{\omega : \xi_{n_j} = 1 \text{ for } j = 1, \ldots, T \text{ and} \\ \xi_m = 0 \text{ if } m \notin \{n_1, \ldots, n_T\},\ m \in [q^N, S-1]\}. \end{aligned}$$

Further set

$$\sigma_n = \frac{\lambda_n}{1 - \lambda_n} = \frac{1}{\log(n/e)} \qquad \text{and} \qquad W_{N,S} = \prod_{q^N \le n < S} (1 - \lambda_n).$$

From the above definitions of $B_{n_1,\ldots,n_T}$, σ_n and $W_{N,S}$, it follows that

$$P(B_{n_1,\ldots,n_T}) = \sigma_{n_1} \cdots \sigma_{n_T} W_{N,S}. \tag{6.3}$$

Let us now introduce the intervals

$$\mathcal{J}_a = [q^N + aq, q^N + (a+1)q - 1] \qquad (a = 0, 1, \ldots, R-1),$$

so that

$$[q^N, S] = \bigcup_{a=0}^{q^N - q^{N-1} - 1} \mathcal{J}_a.$$

Given $n_1, \ldots, n_T$, let $\mathcal{J}_{M_1}, \ldots, \mathcal{J}_{M_H}$ be those intervals which contain at least two elements (say, $\mathcal{J}_{M_j}$ contains $k_j \ge 2$ elements) from the set $\{n_1, \ldots, n_T\}$.

Then it follows from (6.2) that

$$\sum_{\sum_{j=1}^{H} k_j \geq cq^N/N^2} P(B_{n_1,\ldots,n_T}) < \frac{c_1}{N^2}.$$

Let us now consider those $n_1, \ldots, n_T$ for which $\sum_{j=1}^{H} k_j < cq^N/N^2$. Consider those elements amongst $n_1, \ldots, n_T$ which have $k_j \geq 2$ fixed elements in $\mathcal{J}_{M_j}$ $(j = 1, \ldots, H)$ and exactly one element in the intervals $\mathcal{J}_{a_1}, \ldots, \mathcal{J}_{a_U}$.

Define the quantities L and U by

$$\sum_{j=1}^{H} k_j = L \quad \text{and} \quad U = T - L.$$

Here $0 \leq a_1 < \cdots < a_U \leq R - 1$ are such that

$$\{a_1, \ldots, a_U\} \cap \{M_1, \ldots, M_H\} = \emptyset.$$

We shall denote that particular set of $n_1, \ldots, n_T$ by $D(a_1, \ldots, a_{U_N})$, that is a set that contains exactly q^U disjoint sets.

Since, for $j = 1, \ldots, q-1$, we have

$$0 \leq \sigma_n - \sigma_{n+j} \leq \frac{c}{n \log^2 n} \leq \frac{c_1}{q^N N^2},$$

it follows from (6.3) that

$$P(B_{n_1,\ldots,n_T}) = \prod_{j=1}^{U} \sigma_{q^N + a_j q} \cdot \prod_{\ell=1}^{H} \sigma_{q^N + M_\ell q} \cdot W_{N,S} \left(1 + O\left(\frac{1}{N^2 q^N}\right)\right)^T.$$

Since $T/N^2 q^N \to 0$ as $N \to \infty$, it follows that

$$\left(1 + O\left(\frac{1}{N^2 q^N}\right)\right)^T = 1 + o(1) \quad \text{as} \quad N \to \infty.$$

All this means that

$$P(B_{n_1,\ldots,n_T}) = (1 + o(1)) P(B_{n'_1,\ldots,n'_T}) \tag{6.4}$$

if $n_1, \ldots, n_T$ and $n'_1, \ldots, n'_T$ belong to the same set $D(a_1, \ldots, a_U)$.

For a given outcome ω, we now consider

$$\eta_q^{(N,S)}(\omega) := \text{res}_q(n_1) \ldots \text{res}_q(n_T) \tag{6.5}$$

and we let $F_\beta(\eta_q^{(N,S)}(\omega))$ be the number of occurrences of the word β as a subword in $\eta_q^{(N,S)}(\omega)$. Now, setting

$$Z_{N,S} := \left\{ \omega : |Y_{N,S} - E(Y_{N,S})| > N \sqrt{\frac{q^{N+1}}{\log q^{N+1}}} \right\}$$

and

$$V_{N,S} := \left\{ \omega : \theta_{N,S} > c_1 \frac{q^N}{N^2} \right\},$$

it follows from (6.1) and (6.2) that

$$P(Z_{N,S}) + P(V_{N,S}) < \frac{c}{N^2}.$$

Now, assume that $\omega \notin Z_{N,S} \cup V_{N,S}$. Then, recall definition (6.5) and set

$$S_\beta(n_1, \dots, n_T) = \#\{\eta_q^{(N,S)} = \gamma_1 \beta \gamma_2 : \gamma_1, \gamma_2 \in A_q^*\}.$$

In the set $D(n_1, \dots, n_T)$, the elements from $\mathcal{J}_{a_j}$ can be written as $q^N + qa_j + \ell_j$, where $\ell_j \in A_q$ $(j = 1, \dots, U)$. Furthermore, write each integer $\mu \le q^U - 1$ as

$$\mu = \ell_1 + \ell_2 q + \cdots + \ell_U q^{U-1}.$$

Then we get

$$E_\beta(\mu) = \#\{(\gamma_1, \gamma_2) \in A_q^* \times A_q^* : \ell_1 \dots \ell_U = \gamma_1 \beta \gamma_2\}.$$

Using Lemma 2, we then obtain that

$$|S_\beta(n_1, \dots, n_T) - E_\beta(\mu)| \le 2k(H+1).$$

Now $D(a_1, \dots, a_U)$ is characterized by choosing all possible values $\ell_1 \dots \ell_U \in A_q^U$. Hence, letting $\delta_N = 1/N$, we can apply Lemma 1 and obtain that the number of those $\ell_1 \dots \ell_U$ for which

$$\left| E_\beta(\mu) - \frac{U}{q^k} \right| > U\delta_N$$

is less than $\dfrac{cq^N}{U\delta_N^2}$. Hence, from (6.4), we obtain that

(6.6)
$$\sum_{\substack{(n_1,\dots,n_T) \in D(a_1,\dots,a_U) \\ \left| E_\beta(\mu) - \frac{U}{q^k} \right| > U\delta_N}} P(B_{n_1,\dots,n_T}) < \frac{c_1}{E(Y_{N,S})\delta_N^2} \sum_{(n_1,\dots,n_T) \in D(a_1,\dots,a_U)} P(B_{n_1,\dots,n_T}).$$

Now, summing the inequality (6.6) over all possible values of $n_1, \dots, n_T$ for which $\omega \notin Z_{N,S} \cup V_{N,S}$, the "new" right hand side of (6.6) is then no larger than $\dfrac{c_1}{E(Y_{N,S})} N^2$.

Collecting the above inequalities, we obtain that if

$$K_{N,S} := \left\{ \omega : \omega \notin Z_{N,S},\ \left| E_\beta(\eta_q^{(N,S)}) - \frac{T}{q^k} \right| \ge \frac{2T}{N} \right\},$$

then

$$P(K_{N,S}) < \frac{c}{N^2}.$$

Hence, if we let

$$S_j = q^N + q^N \cdot \frac{j}{\lfloor \log N \rfloor} \qquad (j = 1, \ldots, m_N),$$

where $m_N = (q^2 - q)\lfloor \log N \rfloor$, we then have

$$P\left(\bigcup_{j=1}^{m_N} \left(K_{N,S_j} \cup Z_{N,S_j} \cup V_{N,S_j} \right) \right) < \frac{c \log N}{N^2}. \tag{6.7}$$

Let Q be the set of those ω which belong to infinitely many of the sets $K_{N,S_j} \cup Z_{N,S_j} \cup V_{N,S_j}$. Now, summing (6.7) on $N = 1, \ldots, \infty$, we obtain a finite sum. We may therefore apply the Borel-Cantelli Lemma (Lemma 3) and conclude that $P(Q) = 0$.

Now let $M_1 < M_2 < \cdots$ be the sequence of integers which are the members of the set $\left\{ q^N + \frac{a}{\lfloor \log N \rfloor} q^{N-1} : a = 0, \ldots, m_N,\ N = 3, 4, \ldots \right\}$, and let $\omega \notin Q$. Then, regarding the sequence

$$\xi_R = \mathrm{res}_q(m_1) \ldots \mathrm{res}_q(m_R),$$

we have that

$$\frac{F_\beta(\xi_{M_j}(\omega))}{M_j} \to \frac{1}{q^k} \qquad (j \to \infty). \tag{6.8}$$

Since $1 \leq \frac{M_{j+1}}{M_j} \to 1$ as $j \to \infty$, it follows from (6.8) that the relation

$$\frac{F_\beta(\xi_n(\omega))}{n} \to \frac{1}{q^k} \qquad (n \to \infty)$$

also holds. Since this assertion is true for every finite word β, the proof of the theorem is complete.

References

[1] P.T. Bateman and R.A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363-367.

[2] P.T. Bateman and R.A. Horn, *Primes represented by irreducible polynomials in one variable*, Proc. Sympos. Pure Math. **8** (1965), 119-135.

[3] J.M. De Koninck and I. Kátai, *Construction of normal numbers by classified prime divisors of integers*, Functiones et Approximatio **45.2** (2011), 231–253.

[4] J.M. De Koninck and I. Kátai, *Construction of normal numbers by classified prime divisors of integers II*, Functiones et Approximatio (to appear).

[5] J.M. De Koninck and I. Kátai, *Normal numbers created from primes and polynomials*, Uniform Distribution Theory **7** (2012), no.2, 1–20.

[6] J.M. De Koninck and I. Kátai, *Some new methods for constructing normal numbers*, Annales des Sciences Mathématiques du Québec (to appear).

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

[8] J. Galambos, *Introductory Probability Theory*, Marcel Dekker, New York, 1984.

[9] A. Schinzel and W. Sierpinski, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185-208; Corrigendum: ibid. **5** (1959), 259.

[10] D. Shiu, *Strings of congruent primes*, J. London Math. Soc. (2) **61** (2000), no.2, 359-363.

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 11 septembre 2012; fichier: normal-primes-2012-fea.tex