ACTA ARITHMETICA
LXXII.2 (1995)

On the distribution of subsets of primes in the prime factorization of integers

by

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

1. Introduction and notation. Denote by $\wp$ the set of all prime numbers. Assume that d is a fixed positive integer and that $\wp_0, \wp_1, \ldots, \wp_d$ are disjoint subsets of primes such that

$$\wp_0 \cup \wp_1 \cup \ldots \cup \wp_d = \wp,$$

where $\wp_0$ contains at most finitely many primes (and in fact may be empty).

Let $\pi([a,b])$ be the number of primes belonging to the interval $[a,b]$.

Let $\pi(I|\wp_i) = \#\{p \in \wp_i \cap I\}$, where I is a subset of the integers.

In what follows we assume that

$$\pi([u,u+v]|\wp_i) = \delta_i \pi([u,u+v]) + O\biggl(\frac{u}{(\log u)^{c_1}}\biggr) \tag{1.1}$$

holds uniformly for $2 \le v \le u$, $i = 1, \ldots, d$, where c_1 is a positive constant and $\delta_1, \ldots, \delta_d$ are positive constants such that $\sum_{i=1}^{d} \delta_i = 1$. With the proper rearrangement, we may assume that $\delta_1 \le \ldots \le \delta_d$.

We shall use the notations

$$x_1 = \log x, \quad x_2 = \log\log x, \quad \text{etc.}$$

and

$$t_k(x) = \frac{x_2^{k-1}}{(k-1)!}. \tag{1.2}$$

Writing $f(x) \asymp g(x)$ means that the two functions $f(x)$ and $g(x)$ are of the same order as $x \to \infty$.

Further, denote by $\omega(n) = \sum_{p|n} 1$ the number of distinct prime factors of n, by $P(n)$ the largest prime factor of n and by $p(n)$ the smallest prime factor of n.

Research of the first author partially supported by grants from NSERC of Canada and FCAR of Québec.

Research of the second author partially supported by the Hungarian Research Foundation (OTKA).

In what follows, $p_1, p_2, \ldots$ as well as $q_1, q_2, \ldots$ always denote prime numbers.

An expression of the form $i_1 i_2 \ldots i_t$, where $t \geq 1$ and each i_j is one of the numbers $1, \ldots, d$, is called a *word* of length t. We sometimes write $\lambda(\alpha) = t$ to indicate that α is a word of length t. Let $\mathcal{A}_t$ be the set of all words of length t. Define $\mathcal{A}_0$ to be the set containing the *empty word* Λ. Finally, we set

$$\mathcal{A}^* := \bigcup_{t=0}^{\infty} \mathcal{A}_t.$$

We now define the function $H : \mathbb{N} \to \mathcal{A}^*$ as follows. First let $H(1) = \Lambda$. For an arbitrary prime number p and positive integer a, define

$$H(p^a) = \begin{cases} \Lambda & \text{if } p \in \wp_0, \\ j & \text{if } p \in \wp_j. \end{cases}$$

Further, for $n = p_1^{a_1} \ldots p_r^{a_r}$ $(p_1 < \ldots < p_r)$, define

$$H(n) = H(p_1^{a_1}) \ldots H(p_r^{a_r}).$$

Finally, given a word $\alpha = i_1 \ldots i_t$, we set

$$\varrho(\alpha) := \delta_{i_1} \ldots \delta_{i_t}.$$

Let w_x be a function tending to ∞ but satisfying $w_x = O(x_3)$. For an arbitrary number $w \geq 1$, and for each word $\alpha = i_1 \ldots i_k \in \mathcal{A}_k$, define

$$\begin{aligned} \mathcal{N}_k(w) &= \{p_1^{a_1} \ldots p_k^{a_k} :\ w < p_1 < \ldots < p_k\}, \\ \mathcal{N}_k^{(0)}(w) &= \{p_1 \ldots p_k :\ w < p_1 < \ldots < p_k\}, \\ \mathcal{N}_k(w;\alpha) &= \{p_1^{a_1} \ldots p_k^{a_k} :\ w < p_1 < \ldots < p_k,\ H(p_1^{a_1} \ldots p_k^{a_k}) = \alpha\}, \\ \mathcal{N}_k^{(0)}(w;\alpha) &= \{p_1 \ldots p_k :\ w < p_1 < \ldots < p_k,\ H(p_1 \ldots p_k) = \alpha\}, \\ \mathcal{N}_k^{(1)}(w;\alpha) &= \mathcal{N}_k(w;\alpha) \setminus \mathcal{N}_k^{(0)}(w;\alpha), \\ \mathcal{N}_k^{(1)}(w) &= \mathcal{N}_k(w) \setminus \mathcal{N}_k^{(0)}(w). \end{aligned}$$

For each of the above expressions $\mathcal{N}_k(w), \mathcal{N}_k^{(0)}(w), \mathcal{N}_k(w;\alpha), \ldots$, we define the corresponding counting functions $N_k(Y|w)$, $N_k^{(0)}(Y|w)$, $N_k(Y|w;\alpha), \ldots$ which stand for the number of elements $n \leq Y$ which belong to the corresponding set.

Furthermore, when $w = 1$, we shall write $N_k(Y), N_k^{(0)}(Y)$ and $N_k^{(1)}(Y)$ instead of $N_k(Y|1)$, $N_k^{(0)}(Y|1)$ and $N_k^{(1)}(Y|1)$ respectively.

We shall also use the standard functions related to the normal distribution, namely

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2}\, dt. \tag{1.3}$$

For each number $w \geq 2$ and $z \geq 0$, let

$$\varphi_w(z) := \prod_{p \leq w} \left(1 + \frac{z}{p}\right)^{-1}. \tag{1.4}$$

For $s \in \mathbb{C}$, $\Re(s) > 1$, $z > 0$ and $w > 2$, let

$$E(s,z) = \prod_{p > w} \frac{1 + \frac{z}{p^s - 1}}{1 + \frac{z}{p^s}}. \tag{1.5}$$

Also, for each $z \geq 0$, let

$$F(z) := \frac{1}{\Gamma(z)} \prod_{p} \left(1 + \frac{z}{p}\right)\left(1 - \frac{1}{p}\right)^{z}, \tag{1.6}$$

where $\Gamma(z)$ stands for the Gamma function.

Throughout the text, c stands for a positive constant not necessarily the same at each occurrence. On the other hand, the constants $c_1, c_2, \ldots$ appear at specific occasions and keep their original value throughout the whole text.

2. Preliminary results. In this paper, we are proving several results involving the distribution of subsets of primes in the factorization of integers. Many of these results are stated and proved throughout the paper after having been properly motivated by the flow of the material presented. In this section, however, we state two important preliminary results:

THEOREM 1. *Let $c_1 \geq 5$, $\lim_{x \to \infty} w_x = \infty$, $w_x = O(x_3)$, $\sqrt{x} \leq Y \leq x$ and $1 \leq k \leq c_2 x_2$, where c_2 is an arbitrary constant. Assume that α is an arbitrary word belonging to $\mathcal{A}_k$. Then*

$$N_k(Y|w_x;\alpha) = (1 + o(1))\varrho(\alpha) N_k(Y|w_x). \tag{2.1}$$

THEOREM 2. *Assume that the conditions of Theorem 1 hold. Let $A \leq x_2$ with $P(A) \leq w_x$. Then the number of integers $n = An_1 \leq Y$ for which $p(n_1) > w_x$, $H(n_1) = \alpha$, $\omega(n_1) = k$ and $\alpha = i_1 \ldots i_k$, is equal to*

$$(1 + o(1))\varrho(\alpha) \frac{Y}{A \log Y} t_k(Y) \varphi_{w_x}\left(\frac{k-1}{x_2}\right) F\left(\frac{k-1}{x_2}\right), \tag{2.2}$$

where the functions t_k, φ_{w_x} and F are defined by (1.2), (1.4) and (1.6) respectively.

3. The preliminary lemmas

LEMMA 1. *Define*

$$T_k(x|w) = \sum_{\substack{p_1\dots p_k\le x\\ w<p_1<\dots<p_k}} G(p_1)\dots G(p_k), \tag{3.1}$$

where $G(p)=1+t_p$, with t_p a sequence of real numbers, $0<t_p\le 1$, such that $\sum_p t_p/p<\infty$. Then

$$T_k(x|w) = O\Big(\frac{x}{\log x}\cdot\frac{1}{(\log w)^{k/x_2}}\cdot\frac{x_2^{k-1}}{(k-1)!}\Big) \qquad (1\le k\le c_2x_2). \tag{3.2}$$

Proof. Clearly we have

$$T_k(x|w)\le O(\sqrt{x})+\frac{2}{\log x}\sum_{\substack{p_1\dots p_k\le x\\ w<p_1<\dots<p_k}} G(p_1)\dots G(p_k)\log(p_1\dots p_k). \tag{3.3}$$

Denote this last sum by Σ_0. Then we have

$$\Sigma_0\le \sum_{\substack{q_1\dots q_{k-1}p\le x\\ w<q_1<\dots<q_{k-1}}} G(q_1)\dots G(q_{k-1})G(p)\log p.$$

Using the fact that $\sum_{p\le z}G(p)\log p<cz$ for some positive constant c, we obtain

$$\begin{aligned}\Sigma_0 &\le c\sum_{w<q_1<\dots<q_{k-1}<x} G(q_1)\dots G(q_{k-1})\sum_{p\le x/(q_1\dots q_{k-1})} G(p)\log p \\ &\le cx\frac{1}{(k-1)!}\Big(\sum_{w<q<x}\frac{G(q)}{q}\Big)^{k-1}.\end{aligned} \tag{3.4}$$

Now

$$\sum_{w<q<x}\frac{G(q)}{q}\le x_2-\log\log w+\varepsilon_x, \tag{3.5}$$

where $\varepsilon_x\to 0$ as $x\to\infty$. On the other hand,

$$\begin{aligned}(x_2-\log\log w+\varepsilon_x)^{k-1} &= x_2^{k-1}\Big(1-\frac{\log\log w-\varepsilon_x}{x_2}\Big)^{k-1}\\ &\ll x_2^{k-1}e^{-(\log\log w)k/x_2}\le x_2^{k-1}\frac{c}{(\log w)^{k/x_2}},\end{aligned}$$

for some positive constant $c>0$. This last estimate, combined with (3.3), (3.4) and (3.5), proves (3.2).

LEMMA 2. *Let $c_1\ge 5$, $\lim_{x\to\infty}w_x=\infty$, $w_x=O(x_3)$, $\sqrt{x}\le Y\le x$ and $1\le k\le c_2x_2$. Then, writing for short $w=w_x$, the following three estimates hold:*

$$N_k^{(0)}(Y|w) = \frac{Y}{\log Y}t_k(Y)\varphi_w\Big(\frac{k-1}{x_2}\Big)F\Big(\frac{k-1}{x_2}\Big)\Big(1+O\Big(\frac{k}{x_2^2}\Big)\Big), \tag{3.6}$$

$$\begin{aligned}N_k(Y|w) = \frac{Y}{\log Y}t_k(Y)E\Big(1,\frac{k-1}{x_2}\Big)\varphi_w\Big(\frac{k-1}{x_2}\Big)F\Big(\frac{k-1}{x_2}\Big)&\\ \times\Big(1+O\Big(\frac{k}{x_2^2}\Big)\Big)&,\end{aligned} \tag{3.7}$$

$$N_k^{(1)}(Y|w) = O\Big(\frac{1}{w\log w}N_k^{(0)}(Y|w)\Big). \tag{3.8}$$

Proof. Consider the generating function

$$1+\sum_{\substack{n=2\\ p(n)>w}}^{\infty}\frac{z^{\omega(n)}|\mu(n)|}{n^s}=\prod_{p>w}\Big(1+\frac{z}{p^s}\Big)=\zeta^z(s)h(s),$$

where

$$h(s):=\prod_{p\le w}\Big(1+\frac{z}{p^s}\Big)^{-1}\prod_p\Big(1+\frac{z}{p^s}\Big)\Big(1-\frac{1}{p^s}\Big)^z.$$

Using an analytic method successively developed and refined by Sathe, Selberg and Kubilius (see Kubilius [6]), one can prove that

$$S(x):=\sum_{\substack{n\le Y\\ p(n)>w}} z^{\omega(n)}|\mu(n)|=\Big(\frac{h(1)}{\Gamma(z)}+O\Big(\frac{1}{\log Y}\Big)\Big)Y\log^{z-1}Y,$$

from which we easily deduce (3.6).

Similarly, starting with

$$1+\sum_{\substack{n=2\\ p(n)>w}}^{\infty}\frac{z^{\omega(n)}}{n^s}=\prod_{p>w}\Big(1+\frac{z}{p^s-1}\Big)=\prod_{p>w}\Big(1+\frac{z}{p^s}\Big)E(s,z),$$

we obtain (3.7).

Finally, since

$$E(s,z)=\prod_{p>w}\frac{1+\frac{z}{p-1}}{1+\frac{z}{p}}=1+O\Big(\frac{1}{w\log w}\Big),$$

a relation valid for $0\le z<c$, we deduce (3.8). This ends the proof of Lemma 2.

4. The proofs of Theorems 1 and 2. First we prove Theorem 1. Let c_3 be a positive constant, $l_0<l_1<\dots$ be knot-points in the interval $[w_x,Y]$ such that $l_0=w_x$, $l_{j+1}=l_j+l_j/(\log l_j)^{c_3}$ $(j=1,2,\dots)$. We also define l_{-1} by the equation $l_{-1}+l_{-1}/(\log l_{-1})^{c_3}=l_0$. A k-tuple $(u_1,\dots,u_k)$ of knot-points is said to be *feasible* if it satisfies $l_0\le u_1\le\dots\le u_k$ and

$u_1 \dots u_k \le Y$. Further, let $\widetilde{u}_j = [u_j, u_j + \Delta u_j]$. Here $\Delta u_j = l_{k+1} - l_k$ if $u_j = l_k$.

Let $\boldsymbol{u} = (u_1, \dots, u_k)$ be a feasible k-tuple and, given $\alpha = i_1 \dots i_k$, write

$$\pi_k(\boldsymbol{u}|\alpha) := \#\{(p_1, \dots, p_k) : p_j \in \widetilde{u}_j \text{ and } H(p_j) = i_j, \text{ for } j = 1, \dots, k\}.$$

We also define

$$\pi_k(\boldsymbol{u}) = \sum_{\lambda(\alpha)=k} \pi_k(\boldsymbol{u}|\alpha),$$

where the sum runs through all words α of length k.

Since we have assumed (see (1.1)) that

$$\pi(\widetilde{u}_\nu|\wp_j) = \delta_j \pi(\widetilde{u}_\nu)\left(1 + O\left(\frac{1}{(\log u_\nu)^{c_1-c_3}}\right)\right),$$

it follows that

$$\frac{1}{S(\boldsymbol{u})} \le \frac{\pi_k(\boldsymbol{u}|\alpha)}{\varrho(\alpha)\pi_k(\boldsymbol{u})} \le S(\boldsymbol{u}), \tag{4.1}$$

where

$$S(\boldsymbol{u}) := \prod_{\nu=1}^{k} \left(1 + \frac{c_4}{(\log u_\nu)^{c_1-c_3}}\right), \tag{4.2}$$

and $c_4 > 0$ is a large constant.

Let $c_5 > 0$ be another constant which is to be determined implicitly by (4.5).

The feasible $\boldsymbol{u}$'s are subdivided into three classes, $\mathcal{B}_0$, $\mathcal{B}_1$ and $\mathcal{B}_2$, as follows:

- $\boldsymbol{u} \in \mathcal{B}_0$ if there exists at least one ν for which

$$u_{\nu+1} - u_\nu \le \frac{u_\nu}{(\log u_\nu)^{c_5}}, \tag{4.3}$$

- $\boldsymbol{u} \in \mathcal{B}_1$ if $\boldsymbol{u} \notin \mathcal{B}_0$ and

$$(u_1 + \Delta u_1) \dots (u_k + \Delta u_k) > Y, \tag{4.4}$$

- $\mathcal{B}_2$ contains all the other $\boldsymbol{u}$'s.

First observe that if $\boldsymbol{u} \notin \mathcal{B}_0$ and s_t denotes the number of $u_\nu \in [e^t, e^{t+1}]$, then $s_t \ll t^{c_5}$, and consequently, if we set $t_0 = [\log l_0]$, we have

$$\log S(\boldsymbol{u}) \ll \sum_{t \ge t_0} \frac{t^{c_5}}{t^{c_1-c_3}} \ll \frac{1}{t_0^{c_1-c_3-c_5-1}} = O\left(\frac{1}{\log w_x}\right),$$

provided

$$c_5 + c_3 + 2 \le c_1. \tag{4.5}$$

We have thus proved that

$$S(\boldsymbol{u}) = 1 + O\left(\frac{1}{\log w_x}\right) \quad \text{for } \boldsymbol{u} \notin \mathcal{B}_0. \tag{4.6}$$

From (4.6) and (4.1), it follows that

$$\sum_{\boldsymbol{u} \in \mathcal{B}_\xi} \pi_k(\boldsymbol{u}|\alpha) = \varrho(\alpha)\left(1 + O\left(\frac{1}{\log w_x}\right)\right) \sum_{\boldsymbol{u} \in \mathcal{B}_\xi} \pi_k(\boldsymbol{u}) \qquad (\xi = 1, 2). \tag{4.7}$$

On the other hand, it is clear that

$$\sum_{\boldsymbol{u} \in \mathcal{B}_2} \pi_k(\boldsymbol{u}|\alpha) \le N_k^{(0)}(Y|w_x; \alpha)$$
$$\le \sum_{\boldsymbol{u} \in \mathcal{B}_1} \pi_k(\boldsymbol{u}|\alpha) + \sum_{\boldsymbol{u} \in \mathcal{B}_2} \pi_k(\boldsymbol{u}|\alpha) + \sum_{\boldsymbol{u} \in \mathcal{B}_0} \pi_k(\boldsymbol{u}|\alpha).$$

We now proceed to estimate

$$\Sigma_{1,\alpha} := \sum_{\boldsymbol{u} \in \mathcal{B}_0} \pi_k(\boldsymbol{u}|\alpha).$$

Clearly, because of (4.1), we have

$$\Sigma_{1,\alpha} \le \varrho(\alpha) \sum_{\boldsymbol{u} \in \mathcal{B}_0} \pi_k(\boldsymbol{u}) S(\boldsymbol{u}) = \varrho(\alpha)(\Sigma_{1,1} + \Sigma_{1,2}), \tag{4.8}$$

where in $\Sigma_{1,1}$ we sum over those $\boldsymbol{u} \in \mathcal{B}_0$ for which $(u_1 + \Delta u_1) \dots (u_k + \Delta u_k) \le Y$, and in $\Sigma_{1,2}$ we sum over the other $\boldsymbol{u}$'s.

Now define

$$G(p) = 1 + \frac{2c_4}{(\log p)^{c_1-c_3}}.$$

We then have

$$\Sigma_{1,1} \le \sideset{}{^*}\sum_{p_1 \dots p_k \le Y} G(p_1 \dots p_k), \tag{4.9}$$

where the asterisk in the sum indicates that $0 < p_{\nu+1} - p_\nu < 2p_\nu/(\log p_\nu)^{c_5}$ is satisfied for at least one $\nu \in [1, k]$.

In order to estimate $\Sigma_{1,2}$, we replace $\boldsymbol{u}$ by $\boldsymbol{u}' = (u'_1, \dots, u'_k)$ where u'_l is the left neighbour of u_l among the knot-points. If l_0 occurs among the u_t's, then it is simply shifted into l_{-1}. Note that it is clear that $l_{-1} \ge l_0/2$.

By construction, we have

$$\pi(\widetilde{u}'_\nu) = \pi(\widetilde{u}_\nu)\left(1 + O\left(\frac{1}{(\log u_\nu)^2}\right)\right),$$

say, and since $u'_1 \dots u'_k \le Y$, it follows that

$$\Sigma_{1,2} \ll \sideset{}{^{**}}\sum_{p_1 \dots p_k \le Y} G'(p_1 \dots p_k) = \Sigma_A, \tag{4.10}$$

say, where the double asterisk in the sum indicates that we sum over those

$p_1, \dots, p_k$ for which $0 < p_{\nu+1} - p_\nu < 2p_\nu/(\log p_\nu)^{c_5}$ holds for at least one ν, $l_{-1} \le p_1 < \dots < p_k$, and where $G' = 1 + f_p$, with $f_p = 1/\log p$.

The sums (4.9) and (4.10) being similar, we only need to find an upper bound for Σ_A. First we set $g(p) = 2p/(\log p)^{c_5}$.

Clearly we have

$$(4.11)\quad \Sigma_A \le 4 \sum_{\substack{pq<Y \\ p<q<p+g(p)}} \sum_{\substack{q_1\dots q_{k-2} \le Y/qp \\ l_{-1}\le q_1<\dots<q_{k-2}}} G'(q_1 \dots q_{k-2}) = 4\sum_{p,q} \sum_{q_1,\dots,q_{k-2}},$$

say. We consider the cases $p > Y^{1/10}$ and $p \le Y^{1/10}$ separately. For the first case, since $G(p) \le 2$ and

$$\sum_{p<q<p+g(p)} \frac{1}{q} \ll \frac{1}{(\log p)^{c_5+1}},$$

we have

$$(4.12)\quad \sum_{\substack{p>Y^{1/10} \\ p<q<p+g(p)}} \sum_{q_1,\dots,q_{k-2}} \le 2^{k-2} Y \sum_{p>Y^{1/10}} \frac{1}{p} \sum_{p<q<p+g(p)} \frac{1}{q}$$
$$\ll 2^k Y \sum_{p>Y^{1/10}} \frac{1}{p(\log p)^{c_5+1}}$$
$$\ll 2^k Y \int_{Y^{1/10}}^{\infty} \frac{dt}{t(\log t)^{c_5+2}} \ll \frac{Y}{(\log x)^{c_5}},$$

because $2^k = O(\log x)$ and since $\log Y \asymp \log x$.

For the second case, we use the inequality (3.2) of Lemma 1 and obtain

$$(4.13)\quad \sum_{\substack{p\le Y^{1/10} \\ p<q<p+g(p)}} \sum_{q_1,\dots,q_{k-2}}$$
$$\ll \sum_{p\le Y^{1/10}} \frac{Y}{x_1} \cdot \frac{1}{(k-3)!} \Bigg(\sum_{l_{-1}<q<x} \frac{1+f_p}{q} \Bigg)^{k-3} \sum_{p<q<p+g(p)} \frac{1}{pq}$$
$$\ll \frac{Y}{x_1} \Bigg(\log \frac{\log x}{\log l_{-1}} + o(1) \Bigg)^{k-1} \frac{1}{(k-1)!} \sum_{p<q<p+g(p)} \frac{1}{pq}$$
$$\ll \frac{Y}{x_1} \cdot \frac{x_2^{k-3}}{(k-3)!} \cdot \frac{1}{(\log w_x)^{k/x_2}} \cdot \frac{1}{(\log w_x)^{c_5}} \sum_{p<q<p+g(p)} \frac{1}{pq}$$
$$= O\Bigg(\frac{Y}{x_1} \cdot \frac{1}{(\log w_x)^{k/x_2+c_5}} t_{k-2}(x) \Bigg)$$

as $x \to \infty$. Thus in view of (4.8) and the estimates (4.11)–(4.13), as well as Lemma 2, we have proved that, for every word α of length k,

$$\Sigma_{1,\alpha} = o(1)\varrho(\alpha) N_k^{(0)}(Y|w_x)$$

and in particular that

$$\sum_{\boldsymbol{u}\in\mathcal{B}_0} \pi_k(\boldsymbol{u}) = \sum_{\alpha} \Sigma_{1,\alpha} = o(1) N_k^{(0)}(Y|w_x).$$

If $\boldsymbol{u} \in \mathcal{B}_1$, then, using (4.6), we have

$$Y < (u_1 + \Delta u_1)\dots(u_k + \Delta u_k) \le YS(\boldsymbol{u}) \le Y + O\Bigg(\frac{Y}{\log w_x}\Bigg),$$

and

$$u_1 \dots u_k \ge Y - O\Bigg(\frac{Y}{\log w_x}\Bigg).$$

Thus, with a suitable large number B, we have, using Lemma 2,

$$(4.14)\quad \sum_{\boldsymbol{u}\in\mathcal{B}_1} \pi_k(\boldsymbol{u}|\alpha) \le \varrho(\alpha)\Bigg(1 + O\Bigg(\frac{1}{\log w_x}\Bigg)\Bigg)$$
$$\times \Bigg\{ N_k^{(0)}\Bigg(Y + \frac{YB}{w_x^2}\Bigg| w_x\Bigg) - N_k^{(0)}\Bigg(Y - \frac{YB}{w_x^2}\Bigg| w_x\Bigg)\Bigg\}$$
$$\ll \frac{1}{w_x^2} \varrho(\alpha) N_k^{(0)}(Y|w_x) = o(1)\varrho(\alpha) N_k^{(0)}(Y|w_x).$$

Hence

$$\sum_{\boldsymbol{u}\in\mathcal{B}_1} \pi_k(\boldsymbol{u}) \ll o(1) N_k^{(0)}(Y|w_x).$$

We have therefore proved that

$$(4.15)\qquad N_k^{(0)}(Y|w_x;\alpha) = (1 + o(1))\varrho(\alpha) N_k^{(0)}(Y|w_x).$$

We now proceed to estimate $N_k^{(1)}(Y|w_x;\alpha)$, which, as we may recall from the definition given in Section 1, represents the number of positive integers $n = p_1^{a_1} \dots p_k^{a_k} \le Y$, where $w_x < p_1 < \dots < p_k$, such that $H(n) = \alpha$ and have at least one $a_i > 1$.

We write such an n as $n = n_1 n_2$, where n_1 stands for the square-full part of n and n_2 stands for the square-free part of n. Note that we have $\varrho(\alpha) \ge \delta_1^k$. Observe first that we can omit all those integers n for which the corresponding $n_1 \ge x_1^c$, where c is a large constant depending on δ_1: the reason is that their contribution to $N_k^{(1)}(Y|w_x;\alpha)$ is less than $\varrho(\alpha)Y/x_1^2$. We can thus assume that $n_1 \ll (\log x)^c$ for some large constant $c > 0$.

For each n_1, consider those n for which the square-full part is n_1. Let $H(n_2) = \alpha_{n_1}$. Thus, using (4.15), we may write

$$\begin{aligned} N_k^{(1)}(Y|w_x;\alpha) &\le \sum_{\substack{1<n_1<x_1^c \\ n_1 \text{ square-full} \\ p(n_1)>w_x}} N_{k-\omega(n_1)}^{(0)}\left(\frac{Y}{n_1}\middle| w_x;\alpha_{n_1}\right) \\ &\ll \sum_{n_1} \varrho(\alpha_{n_1}) N_{k-\omega(n_1)}^{(0)}\left(\frac{Y}{n_1}\middle| w_x\right) \\ &\ll \varrho(\alpha) \sum_{n_1} \left(\frac{1}{\delta_1}\right)^{\omega(n_1)} N_{k-\omega(n_1)}^{(0)}\left(\frac{Y}{n_1}\middle| w_x\right). \end{aligned}$$

Let us first assume that $k \le x_2$. In this case, $N_{k-\omega(n_1)}^{(0)}(\frac{Y}{n_1}|w_x)$ is essentially monotonic in k, that is,

$$N_{k-\omega(n_1)}^{(0)}\left(\frac{Y}{n_1}\middle| w_x\right) \ll \frac{1}{n_1} N_k^{(0)}(Y|w_x),$$

and therefore

$$N_k^{(1)}(Y|w_x;\alpha) \ll \varrho(\alpha)\left(\sum_{n_1}\left(\frac{1}{\delta_1}\right)^{\omega(n_1)}\frac{1}{n_1}\right) N_k^{(0)}(Y|w_x). \tag{4.16}$$

But since

$$\begin{aligned} \sum_{n_1}\left(\frac{1}{\delta_1}\right)^{\omega(n_1)}\frac{1}{n_1} &= \prod_{p>w_x}\left(1+\frac{1}{\delta_1}\left(\frac{1}{p^2}+\frac{1}{p^3}+\dots\right)\right) \\ &< \exp\left\{\frac{2}{\delta_1}\sum_{p>w_x}\frac{1}{p^2}\right\} - 1 < \frac{4}{\delta_1 w_x \log w_x}, \end{aligned} \tag{4.17}$$

it follows that

$$N_k^{(1)}(Y|w_x;\alpha) \ll \frac{1}{w_x\log w_x} N_k^{(0)}(Y|w_x;\alpha). \tag{4.18}$$

It remains to consider the case where $x_2 < k \le c_2x_2$. In this case, $N_{k-\omega(n_1)}^{(0)}(\frac{Y}{n_1}|w_x)$ is essentially decreasing in k. Hence we proceed as follows. We have

$$N_{k-\omega(n_1)}^{(0)}\left(\frac{Y}{n_1}\middle| w_x\right) \ll \frac{1}{n_1} N_{k-\omega(n_1)}^{(0)}(Y|w_x) \ll \frac{1}{n_1} N_k^{(0)}(Y|w_x)\frac{t_{k-\omega(n_1)}(Y)}{t_k(Y)}.$$

Since

$$\frac{t_{k-\omega(n_1)}(Y)}{t_k(Y)} \ll \prod_{j=1}^{\omega(n_1)}\left(\frac{k-j}{x_2}\right) \le \left(\frac{k-1}{x_2}\right)^{\omega(n_1)} \le c_2^{\omega(n_1)},$$

it follows that

$$N_{k-\omega(n_1)}^{(0)}\left(\frac{Y}{n_1}\middle| w_x\right) \ll \varrho(\alpha)\left(\sum_{n_1>1}\left(\frac{c_2}{\delta_1}\right)^{\omega(n_1)}\frac{1}{n_1}\right) N_k^{(0)}(Y|w_x).$$

But, similarly to (4.17), one can show that

$$\sum_{n_1>1}\left(\frac{c_2}{\delta_1}\right)^{\omega(n_1)}\frac{1}{n_1} = O\left(\frac{1}{w_x\log w_x}\right).$$

Thus (4.18) still holds in the case $x_2 < k \le c_2x_2$.

We have thus proved that, if $1 \le k \le c_2x_2$ and if α is an arbitrary sequence of length k, then

$$N_k^{(1)}(Y|w_x;\alpha) \ll \frac{\varrho(\alpha)}{w_x\log w_x} N_k^{(0)}(Y|w_x). \tag{4.19}$$

A consequence of this result is that $N_k^{(1)}(Y|w_x;\alpha) = o(1)N_k^{(0)}(Y|w_x;\alpha)$ in the whole range $1 \le k \le c_2x_2$, a result which, combined with (4.15), ends the proof of Theorem 1.

Theorem 2 follows easily by taking into consideration (4.15), (4.19) and Lemma 2.

5. Immediate applications. Theorems 1 and 2 have a wide range of applications in number theory. An important one will be treated extensively in Section 7. Nevertheless, here we mention two classical situations where the results of Theorem 1 and of Theorem 2 can be applied.

Congruence classes. Let $D > 1$ be a fixed integer. Subdivide the set of primes $\wp$ into congruence classes mod D, that is, in $d = \varphi(D)$ distinct classes, where φ stands for the Euler function. We have $\wp_0 = \{p : p\,|\,D\}$. Then $\varrho(i) = 1/d$ for each $i \ne 0$, and $\varrho(\alpha) = 1/d^{\lambda(\alpha)}$, where $\lambda(\alpha)$ denotes the length of the word α.

Distribution of primes in special sequences. Let the interval $[0,1]$ be subdivided into disjoint intervals $I_1,\dots,I_d$ of length $|I_\nu| = \delta_\nu$. Let η be an irrational number and set

$$\wp_\nu = \{p \in \wp : p\eta - [p\eta] \in I_\nu\} \qquad (\nu = 1,\dots,d).$$

Assume that η is a number for which the corresponding $\wp_\nu$'s satisfy

$$\pi([u,v]|\wp_\nu) = \delta_\nu\pi([u,v]) + O\left(\frac{u}{(\log u)^{c_1}}\right) \qquad (\nu = 1,\dots,d),$$

for every fixed large number $c_1 > 0$. It is a classical result of I. M. Vinogradov that such a relation holds for almost all irrational numbers η.

6. The main results

THEOREM 3. *Let c_6 be an arbitrary positive constant and assume that $c_1 \geq 5$. Set*

$$P = P_{w,y} := \sum_{w<p<y} \frac{1}{p} \tag{6.1}$$

and assume that as $w \to \infty$, we have $y = y(w) \to \infty$ so that $P_{w,y} \to \infty$. Then, uniformly for $1 \leq k \leq c_6 P_{w,y}$, and uniformly for $\alpha \in \mathcal{A}_k$, we have

$$\frac{1}{\varrho(\alpha)} \sum_{\substack{w<p_1<\ldots<p_k<y \\ H(p_1\ldots p_k)=\alpha}} \frac{1}{p_1\ldots p_k} = (1+o_w(1)) \sum_{w<p_1<\ldots<p_k<y} \frac{1}{p_1\ldots p_k}.$$

Furthermore,

$$\frac{1}{\varrho(\alpha)} \sum_{\substack{w<p_1<\ldots<p_k<y \\ H(p_1^{a_1}\ldots p_k^{a_k})=\alpha \\ \max(a_1,\ldots,a_k)>1}} \frac{1}{p_1^{a_1}\ldots p_k^{a_k}} = o_w(1) \sum_{w<p_1<\ldots<p_k<y} \frac{1}{p_1\ldots p_k}.$$

Proof of Theorem 3. The proof is very similar to that of Theorem 1. Let

$$S_k(\alpha) = \sum_{\substack{w<p_1<\ldots<p_k<y \\ H(p_1\ldots p_k)=\alpha}} \frac{1}{p_1\ldots p_k}, \qquad S_k = \sum_{w<p_1<\ldots<p_k<y} \frac{1}{p_1\ldots p_k}.$$

Divide the interval $[w,y]$ by knot-points $l_0 < \ldots < l_t$, where $l_0 = w$, $l_{i+1} - l_i = l_i/(\log l_i)^{c_7}$, for some constant $c_7 > 0$. For an arbitrary k-tuple of subintervals $\widetilde{u}_\nu = [u_\nu, u_\nu + \Delta u_\nu]$ $(\nu = 1,\ldots,k)$, $u_1 \leq \ldots \leq u_k$, let

$$\mathcal{L}(\boldsymbol{u}) = \sum_{p_\nu \in \widetilde{u}_\nu} \frac{1}{p_1\ldots p_k} \quad \text{and} \quad \mathcal{L}(\boldsymbol{u}|\alpha) = \sum_{\substack{p_\nu \in \widetilde{u}_\nu \\ H(p_1\ldots p_k)=\alpha}} \frac{1}{p_1\ldots p_k},$$

where $\boldsymbol{u} = (u_1,\ldots,u_k)$. Since

$$u_1\ldots u_k \leq p_1\ldots p_k \leq (u_1\ldots u_k)\prod_{\nu=1}^{k}\left(1+\frac{\Delta u_\nu}{u_\nu}\right),$$

it follows that

$$\frac{1}{S(\boldsymbol{u})} \leq \frac{\mathcal{L}(\boldsymbol{u})}{\varrho(\alpha)\mathcal{L}(\boldsymbol{u}|\alpha)} \leq S(\boldsymbol{u}),$$

where

$$S(\boldsymbol{u}) := \prod_{\nu=1}^{k}\left(1+\frac{c}{(\log u_\nu)^{c_7}}\right),$$

for some large constant $c > 0$.

We shall say that $\boldsymbol{u}$ is *well spaced* if $S(\boldsymbol{u}) \leq 1+\varepsilon$, where $\varepsilon > 0$ is an arbitrary but fixed positive number. Hence, if $\boldsymbol{u}$ is not well spaced, it means that there exists at least one couple of primes $p_\nu, p_{\nu+1}$ such that

$$p_\nu < p_{\nu+1} < p_\nu + g(p_\nu),$$

where $g(p) = p/(\log p)^{c_8}$, with a positive constant c_8. We shall see that the main contribution to the sum $S_k(\alpha)$ comes from the well spaced $\boldsymbol{u}$'s.

In order to find an upper bound for the contribution of the badly spaced prime sequences $\{p_1,\ldots,p_k\}$, we subdivide them into classes $J(l_1,t_1,\ldots$ $\ldots,l_r,t_r)$, where the l_ν's and the t_ν's are positive integers such that

$$l_1 < l_1+t_1 < l_2 < l_2+t_2 < \ldots < l_r+t_r \leq k,$$

the subdivision being made according to the following rule: $\{p_1,\ldots,p_k\} \in J(l_1,t_1,\ldots,l_r,t_r)$ if for every ν $(1 \leq \nu \leq r)$,

(a) $p_{l_\nu+j+1} - p_{l_\nu+j} < g(p_{l_\nu+j})$ $(j = 0,1,\ldots,t_\nu-1)$ and

(b) $p_{h+1} - p_h > g(p_h)$, $p_h - p_{h-1} > g(p_{h-1})$ for each $h \notin \bigcup_{\nu=1}^{r}\{l_\nu, l_\nu+1,\ldots,l_\nu+t_\nu\}$.

Further, define

$$P_1 = p_1\ldots p_{l_1-1}, \quad Q_1 = p_{l_1}\ldots p_{l_1+t_1}, \quad \ldots, \quad P_{r+1} = p_{l_r+t_r+1}\ldots p_k.$$

Note that it may happen that P_1 and/or P_{r+1} are empty. Then set

$$U = P_1\ldots P_{r+1}, \qquad V = Q_1\ldots Q_r.$$

Note that the value of V determines its factorization into $Q_1,\ldots,Q_r$. Observe also that the primes occurring in U are well spaced. Furthermore, if V is given, then only one factorization of U exists with the property that $P_1, Q_1, P_2, Q_2, \ldots, P_{r+1}$ contain the primes in increasing order.

Let us now fix both $J(l_1,t_1,\ldots,l_r,t_r)$ and V.

Since $\alpha = H(p_1\ldots p_k) = H(P_1)H(Q_1)H(P_2)H(Q_2)\ldots H(P_{r+1})$, it follows that all the $H(P_\nu) = \beta_\nu$ $(\nu = 1,\ldots,r+1)$ are determined by α. So let us fix $Q_1,\ldots,Q_r$ and consider the sum

$$K_\alpha := \sum_{H(P_\nu)=\beta_\nu}^{*} \frac{1}{P_1\ldots P_{r+1}},$$

where the asterisk indicates that we sum over $\{p_1,\ldots,p_k\} \in J(l_1,t_1,\ldots$ $\ldots,l_r,t_r)$ with the corresponding fixed V. We can compare K_α with

$$K := \sum{}^{*} \frac{1}{P_1\ldots P_{r+1}},$$

where we have dropped the condition $H(P_\nu) = \beta_\nu$ but kept all the others. Since the primes p_i's in $P_1\ldots P_{r+1}$ are well spaced, we have

$$K_\alpha \leq (1+\varepsilon)\varrho(\beta_1)\ldots\varrho(\beta_{r+1})K.$$

Using the fact that

$$\varrho(\alpha) \geq \prod_{i=1}^{r+1} \varrho(\beta_i)(\delta_1)^{\omega(Q_1 \ldots Q_r)},$$

it follows that, denoting by $\mathcal{T}_\alpha$ the contribution of the badly spaced $\boldsymbol{u}$ to the sum, we get, recalling notation (6.1),

$$\begin{aligned}\mathcal{T}_\alpha &\ll \varrho(\alpha) \sum_{l_\nu, t_\nu} \frac{(1/\delta_1)^{\omega(Q_1 \ldots Q_r)}}{(Q_1 \ldots Q_r)(P_1 \ldots P_{r+1})} \\ &\ll \varrho(\alpha) \sum_V \frac{(1/\delta_1)^{\omega(V)}}{V} \sum_{\omega(U)=k-\omega(V)} \frac{1}{U} \\ &\ll \varrho(\alpha) \sum_V \frac{(1/\delta_1)^{\omega(V)}}{V} \cdot \frac{P^{k-\omega(V)}}{(k-\omega(V))!},\end{aligned}$$

where we used the fact that

$$\sum_{w<p_i<y} \frac{1}{p_1 \ldots p_t} \leq \frac{1}{t!} P^t \qquad (t \geq 1).$$

Now since

$$\frac{P^{k-\omega(V)}}{(k-\omega(V))!} = \frac{P^k}{k!} \prod_{j=0}^{\omega(V)} \left(\frac{k-j}{P}\right) \ll \frac{P^k}{k!} \left(\frac{k}{P}\right)^{\omega(V)} \leq \frac{P^k}{k!} c_6^{\omega(V)},$$

we have

$$\mathcal{T}_\alpha \ll \varrho(\alpha) \frac{P^k}{k!} \sum_V \frac{(c_6/\delta_1)^{\omega(V)}}{V},$$

where in $\sum_V$ we sum over all the $V = Q_1 \ldots Q_r$, where the Q_ν's run over those integers all prime factors of which are close to each other in the sense mentioned earlier.

For each fixed r, we have

$$\left(\sum_{Q_1} \frac{1}{Q_1}\right) \cdots \left(\sum_{Q_r} \frac{1}{Q_r}\right) \leq \left(\frac{c_6/\delta_1}{(\log w)^{c_8}}\right)^r,$$

whence it follows that

$$\sum_V \frac{(c/\delta_1)^{\omega(V)}}{V} \ll \frac{1}{(\log w)^{c_4}}.$$

Thus we have

$$\mathcal{T}_\alpha \ll \frac{\varrho(\alpha)}{(\log w)^{c_8}} \cdot \frac{P^k}{k!}.$$

Similarly (and more easily!), one can prove that the contribution of the badly spaced $\{p_1, \ldots, p_k\}$ to S_k can be estimated from above by

$$\sum_\alpha \mathcal{T}_\alpha \ll \frac{1}{(\log w)^{c_8}} \cdot \frac{P^k}{k!}.$$

Now since $S_k \asymp P^k/k!$ in the range $1 \leq k \leq c_2 x_2$, by summing over all the well spaced $\boldsymbol{u}$'s and taking into account the above estimates, the first assertion of the theorem follows. The second assertion can be proved in a similar way; hence we will omit its proof.

Notation. For a given word α, let J_α denote the set of words β such that $\alpha = \beta\gamma$ for some γ, including the word $\beta = \Lambda$. Furthermore, assume that z satisfies $0 \leq z \leq c_9$ for some constant c_9, and let

$$g(z|\alpha, w) := \sum_{\substack{H(A) \in J_\alpha \\ P(A) \leq w}} \frac{z^{\omega(A)}}{A\varrho(H(A))},$$

where the sum runs over all numbers A such that $P(A) \leq w$ and for which $H(A) \in J_\alpha$, including $A = 1$. Let also

$$\kappa(z|\alpha, w) := g(z|\alpha, w)\varphi_w(z), \tag{6.2}$$

where $\varphi_w(z)$ is defined by (1.4). Note also that we shall assume that α is a word of length greater than $\pi(w)$, which implies that $H(A) \in J_\alpha$ has a meaning for each A occurring in the definition of g.

Let $w = w_1$ be fixed for the moment and let $w_2 > w_1$. Then

$$g(z|\alpha, w_2) = \sum_{\substack{H(A_1) \in J_\alpha \\ P(A_1) \leq w_1}} \frac{z^{\omega(A_1)}}{A_1\varrho(H(A_1))} \sum{}^{*} \frac{z^{\omega(A_2)}}{A_2\varrho(H(A_2))}, \tag{6.3}$$

where the asterisk in the inner sum indicates that summation is to be taken over those A_2 for which $H(A_1)H(A_2) \in J_\alpha$ and satisfying $w_1 < P(A_2) \leq P(A_1) \leq w_2$, with $A_2 = 1$ being included.

Assume that the length of α is greater than $\pi(w_2)$. To estimate the inner sum on the right hand side of (6.3), we use Theorem 3; indeed, for

$$t \leq c_{10} \log \frac{\log w_2}{\log w_1} = \tau,$$

we have

$$\sum_{\omega(A_2)=t} \frac{1}{A_2\varrho(H(A_2))} = (1 + o_{w_1}(1)) \sum_{w_1<p_1<\ldots<p_t<w_2} \frac{1}{p_1 \ldots p_t}$$

and thus

$$(6.4)\qquad \sum_{\omega(A_2)\le\tau} \frac{z^{\omega(A_2)}}{A_2\varrho(H(A_2))} = (1+o_{w_1}(1))\sum_{t\le\tau} z^t \sum_{w_1<p_1<\ldots<p_t<w_2} \frac{1}{p_1\ldots p_t}.$$

On the other hand,

$$(6.5)\qquad \sum_{t\le\tau} z^t \sum_{w_1<p_1<\ldots<p_t<w_2} \frac{1}{p_1\ldots p_t}$$

$$= \prod_{w_1<p\le w_2}\left(1+\frac{z}{p}\right) + O\left(\sum_{t>\tau}\frac{z^t}{t!}\left(\sum\frac{1}{p}\right)^t\right).$$

Note that in (6.5) the first term on the right hand side is clearly of order e^{Pz}, where $P=\sum_{w_1<p\le w_2} 1/p$, while the error term is $o(e^{Pz})$ if we assume that $2z<c_{10}$, say. Furthermore,

$$(6.6)\qquad \sum_{\omega(A_2)>\tau}\frac{z^{\omega(A_2)}}{A_2\varrho(H(A_2))} \ll \sum_{\omega(A_2)>\tau}\left(\frac{z}{\delta_1}\right)^{\omega(A_2)}\frac{1}{A_2}$$

$$\ll \sum_{t>\tau}\frac{(z/\delta_1)^t P^t}{t!} = o(1)e^{Pz}$$

if $\tau>c_{11}z$, where c_{11} is a sufficiently large constant. Note that clearly we can also assume that $c_{10}>c_{11}$.

Combining (6.4)–(6.6), we have thus proved that

$$\sum_{A_2}{}^{*}\,\frac{z^{\omega(A_2)}}{A_2\varrho(H(A_2))} = (1+o_{w_1}(1))\prod_{w_1<p\le w_2}\left(1+\frac{z}{p}\right).$$

Hence (6.3) becomes

$$g(z|\alpha,w_2) = (1+o_{w_1}(1))\prod_{w_1<p\le w_2}\left(1+\frac{z}{p}\right)g(z|\alpha,w_1),$$

and consequently,

$$(6.7)\qquad \kappa(z|\alpha,w_2) = (1+o_{w_1}(1))\kappa(z|\alpha,w_1)$$

uniformly in α.

Note that $\kappa(z|\alpha,w)$ depends at most on the first $\pi(w)$ digits of α.

Now given an infinite sequence ξ defined over $\mathcal{A}$, let

$$\kappa(z|\xi,y) := \sum_{\substack{H(A)\in J_\xi\\ P(A)\le y}}\frac{z^{\omega(A)}}{A\varrho(H(A))}\varphi_y(z)$$

and

$$q(z|\xi) := \lim_{y\to\infty}\kappa(z|\xi,y).$$

From (6.7), it is clear that $q(z|\xi)$ exists and furthermore that

$$\tfrac{1}{2}\kappa(z|\xi,w_0)\le q(z|\xi)\le 2\kappa(z|\xi,w_0),$$

provided w_0 is large enough.

For a finite sequence α, we let $\widetilde{\alpha}$ be the infinite sequence $\widetilde{\alpha}=\alpha 1\ldots 1\ldots$

We are now ready to formulate Theorem 4 and deduce its proof mainly by using Theorem 2.

THEOREM 4. *Let k_x be an arbitrary sequence tending to infinity with x. Then, for every k satisfying $k_x\le k\le c_2x_2$ and for every $\alpha\in\mathcal{A}_k$, as $x\to\infty$,*

$$N_k(x|\alpha) = (1+o_x(1))q\left(\frac{k-1}{x_2}\middle|\widetilde{\alpha}\right)\varrho(\alpha)\frac{x}{x_1}t_k(x)F\left(\frac{k-1}{x_2}\right).$$

Proof. Let w_x be the 6-fold iterated logarithm of k_x. Write each n satisfying $H(n)=\alpha$ in the form $n=An_1$, where $P(A)\le w_x$, and $p(n_1)>w_x$. Then clearly

$$(6.8)\qquad N_k(x|\alpha) = \sum_{H(A)\in J_\alpha} N_{k-\omega(A)}\left(\frac{x}{A}\middle| w_x;\gamma_A\right),$$

where γ_A is the word defined implicitly by $\alpha=H(A)\gamma_A$. As in the proof of Theorem 3, we can drop from the sum all the A's for which $A>(\log x)^c$, for some large c, their contribution to the sum being $O(x/(\log x)^c)$. For the other A's, we have, using Theorem 2,

$$(6.9)\qquad N_{k-\omega(A)}\left(\frac{x}{A}\middle| w_x;\gamma_A\right)$$

$$= (1+o_{w_x}(1))\frac{\varrho(\gamma_A)x}{Ax_1}t_{k-\omega(A)}(x)\varphi_{w_x}\left(\frac{k-\omega(A)}{x_2}\right)F\left(\frac{k-\omega(A)}{x_2}\right).$$

Since $\omega(A)$ is small, one can write that

$$t_{k-\omega(A)}(x) = \left(\frac{k-1}{x_2}\right)^{\omega(A)}(1+o_x(1))t_k(x),$$

and since the functions φ_w and F are continuous, (6.9) may be written as

$$(6.10)\qquad N_{k-\omega(A)}\left(\frac{x}{A}\middle| w_x;\gamma_A\right)$$

$$= (1+o_{w_x}(1))\frac{x}{x_1}\cdot\frac{\varrho(\gamma_A)\left(\frac{k-1}{x_2}\right)^{\omega(A)}}{A}t_k(x)\varphi_{w_x}\left(\frac{k-1}{x_2}\right)F\left(\frac{k-1}{x_2}\right).$$

Using (6.10) in (6.8) and the fact that $\varrho(\gamma_A)=\varrho(\alpha)/\varrho(H(A))$, we have

$$N_k(x|\alpha) = (1+o_{w_x}(1))\varrho(\alpha)\frac{x}{x_1}t_k(x) \tag{6.11}$$
$$\times\, \varphi_{w_x}\Big(\frac{k-1}{x_2}\Big)F\Big(\frac{k-1}{x_2}\Big)\sum_{H(A)\in J_\alpha}{}^{*}\ \frac{\big(\frac{k-1}{x_2}\big)^{\omega(A)}}{A\varrho(H(A))},$$

where the asterisk in the sum indicates that we sum for A up to $(\log x)^c$, the contribution of the A's larger than $(\log x)^c$ being $o(1)$; this explains why one can, in view of (6.2) and of the definition of $q(z|\xi)$, replace the sum in (6.11) by $g(z|\alpha, w_x)$ and thereafter $g(z|\alpha, w_x)\,\varphi_{w_x}\big(\frac{k-1}{x_2}\big)$ by $\kappa(z|\alpha, w_x)$, thereby completing the proof of Theorem 4.

7. Counting subwords in $H(n)$. Let β be a particular word in $\mathcal{A}^*$. For an arbitrary $\kappa \in \mathcal{A}^*$, we define $u_\beta(\kappa)$ to be the number of occurrences of β as a subword of κ, i.e. the number of possible $\xi \in \mathcal{A}^*$ for which $\kappa = \xi\beta\eta$ for some $\eta \in \mathcal{A}^*$. For short, we sometimes write $u_\beta(n)$ instead of $u_\beta(H(n))$.

Let $\varrho(\beta)$ be defined as in Section 1, i.e. if $\beta = i_1 \dots i_k$, then $\varrho(\beta) = \varrho(i_1)\dots\varrho(i_k) = \delta_{i_1}\dots\delta_{i_k}$.

By using Theorem 2 and some purely probabilistic theorems we can provide asymptotic estimates for

$$M(x,r,l) = M_\beta(x,r,l) := \#\{n \le x : \omega(n) = r,\ u_\beta(n) = l\}$$

for a wide variety of r and l, and also for

$$M(x,l) = M_\beta(x,l) := \#\{n \le x : u_\beta(n) = l\}.$$

We further need to introduce the quantities $m = m(\beta)$ and $\sigma = \sigma(\beta)$, which represent respectively the mean value and the variance of a random variable X: their exact meaning is given later in (8.3).

THEOREM 5. *Let m and σ be as above. Then as $x \to \infty$,*

$$M(x,r,l) = (1+o(1))\frac{x}{x_1}t_r(x)\frac{m}{\sigma\sqrt{l}}\ \phi\Big(\frac{r-ml}{\sigma\sqrt{l}}\Big), \tag{7.1}$$

uniformly for $r - x_2 = O(x_2/x_3)$, and $l - x_2/m = O(x_2/x_3)$, where ϕ is defined by (1.3). *Furthermore,*

$$M(x,l) = \frac{1+o(1)}{\sqrt{x_2}}\sqrt{\frac{m}{m+\sigma^2}}\ \phi\Big(\sqrt{\frac{m}{m+\sigma^2}}(x_2 - ml)\Big) \tag{7.2}$$

uniformly for $l - x_2/m = O(x_2/x_3)$. Consequently,

$$\lim_{x\to\infty}\frac{1}{x}\#\Big\{n \le x : \frac{u_\beta(n) - x_2/m}{\sqrt{m(m+\sigma^2)}\sqrt{x_2}} < y\Big\} = \Phi(y), \tag{7.3}$$

where $\Phi(y)$ is defined in (1.3).

The proof of Theorem 5 is given in Section 9.

Remark. Most likely, similar assertions are valid for "diophantinely smooth" subsequences of integers, such as substitutional values of polynomials at integer values, or at prime values, but at this moment we are only able to prove such global theorems.

In order to illustrate the method, we shall consider the distribution of the vectorial

$$(u_{\beta_0}(n), u_{\beta_1}(n+1), \dots, u_{\beta_h}(n+h))$$

(see Theorem 6) and the set of shifted primes (see Theorem 7).

In order to do this we set

$$\tau_\beta(n) := \frac{u_\beta(n) - m(\beta)x_2}{c(\beta)\sqrt{x_2}},$$

where $m(\beta) = 1/m$ and $c(\beta) = \sqrt{m(m+\sigma^2)}$. We shall prove the following results.

THEOREM 6. *Assume that $\beta_0, \beta_1, \dots, \beta_h$ are fixed words belonging to $\mathcal{A}^*$. Then*

$$\lim_{x\to\infty}\frac{1}{x}\#\{n \le x : \tau_{\beta_l}(n+l) < y_l\ (l = 0,1,\dots,h)\}$$
$$= \prod_{l=0}^{h}\lim_{x\to\infty}\frac{1}{x}\#\{n \le x : \tau_{\beta_l}(n+l) < y_l\} = \prod_{l=0}^{h}\Phi(y_l).$$

THEOREM 7. *Let $\beta \in \mathcal{A}^*$ be fixed. Then*

$$\lim_{x\to\infty}\frac{1}{\pi(x)}\#\{p \le x : \tau_\beta(p+1) < y\} = \Phi(y).$$

The proofs of these two theorems are given in Sections 10 and 11 respectively.

8. Auxiliary probabilistic results

LEMMA 3. *Let k be a fixed positive integer and let $\xi_0, \xi_1, \dots$ be a sequence of independent random variables, $X_j = f(\xi_j, \xi_{j+1}, \dots, \xi_{j+k-1})$, where f is a Baire function. Let M denote the mean value. Assume that $MX_j = 0$. Let*

$$\sigma^2 = MX_0^2 + 2\sum_{j=1}^{k-1}MX_0X_j \quad (<\infty),$$

and assume that $\sigma \ne 0$. Then

$$\lim_{n\to\infty}P\Big(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}X_j < z\Big) = \phi(z). \tag{8.1}$$

For the proof, see Theorem 19.2.1 in Ibrakhimov and Linnik [5] or Diamanda [1], [2].

LEMMA 4 (Esseen [3]). *Let $X_1, X_2, \ldots$ be independent identically distributed integer valued random variables for which $MX_j = 0$ and $M|X_j|^\varrho < \infty$, with $\varrho \geq 3$. Assume furthermore that for a suitable l, $P(X_j = l) \times P(X_j = l+1) \neq 0$. Then*

$$P(X_1 + \ldots + X_n = k) = \frac{1}{\sigma\sqrt{n}}\phi(z_{n,k}) + O\Big(\frac{1}{n}\Big),$$

where ϕ is defined in (1.3) *and where*

$$z_{n,k} = \frac{k}{\sigma\sqrt{n}}, \qquad \sigma = MX_j^2.$$

Remark. The condition $P(X_j = l)P(X_j = l+1) > 0$ stands only in order to guarantee that the maximal step between possible consecutive values of X is not larger than 1.

Setting up the problem. Let $\mathcal{A} = \{1, \ldots, d\}$. Let ξ_ν be identically distributed independent random variables, $P(\xi_\nu = j) = \delta_j$ $(j = 1, \ldots, d)$, $\delta_j > 0$, $\sum_{j=1}^d \delta_j = 1$. Note that ξ_ν may be an infinite sequence or a finite one.

Let $\beta = b_1 \ldots b_s$, $\gamma = g_1 \ldots g_{s-1}$ be arbitrary but fixed sequences of length s and $s-1$ over $\mathcal{A}$ respectively. For a random sequence $\xi_1 \ldots \xi_n$, we shall denote by $\Pi_\gamma(r)$ the probability of the event that both of the following conditions are satisfied:

1. $\xi_1 \ldots \xi_{s-1} = \gamma$,

2. the number of l's satisfying $1 \leq l \leq n-s+1$ for which $\xi_l\xi_{l+1}\ldots\xi_{l+s-1} = \beta$ is exactly r.

Further, assume that the independent variables Y_i are distributed as the ξ_ν's. Then for an arbitrary $s-1$ tuple γ, let $\Pi_\gamma(t)$ be the probability of the event that

$$g_1 \ldots g_{s-1}Y_1 \ldots Y_t \tag{8.2}$$

ends with β, that is, that $Y_{t-s+1} \ldots Y_t = \beta$ and that this is the only occurrence of β as a subsequence in (8.2). Further, let η_γ denote the length of the sequence $Y_1 \ldots Y_t$. Then $P(\eta_\gamma = t) = \Pi_\gamma(t)$.

Similarly, for the s-tuple β, let $\sigma_\beta(t)$ be the probability of the event that the random sequence

$$b_2 \ldots b_s Y_1 \ldots Y_t$$

has the same property. Thus, using the notation $\beta = b_1\beta^*$, it is clear that $\sigma_\beta(t) = \Pi_{\beta^*}(t)$ and also that $\sum_{t=1}^\infty \sigma_\beta(t) = 1$. Furthermore, let X be the random variable such that $P(X = t) = \Pi_{\beta^*}(t)$.

Finally, let $\tau_\beta(t)$ be the probability of the event that

$$b_2 \ldots b_s Y_1 \ldots Y_t$$

does not contain β as a subword.

It is clear that

$$\tau_\beta(t) = \sigma_\beta(t) + \sigma_\beta(t+1) + \sigma_\beta(t+2) + \ldots,$$

and hence that

$$\tau_\beta(t) = P(X > t).$$

For the random sequence $\xi_1 \ldots \xi_n$ starting with γ, let $t_1 < \ldots < t_r$ be the indices of the last digits of occurrences of the word β. Then clearly $t_1, t_2 - t_1, \ldots, t_r - t_{r-1}, n - t_r$ are independent random variables, where t_1 is distributed as η_γ and the $(t_{j+1} - t_j)$'s are distributed as independent copies of X. Consequently, denoting by $\Pi_\gamma(r,n)$ the probability that $\eta_\gamma + X_1 + \ldots + X_r \leq n$ and that $X_{r+1} > n - (\eta_\gamma + X_1 + \ldots + X_r)$, we have

$$\Pi_\gamma(r,n) = \sum_{\substack{u,v \geq 0 \\ u+v \leq n}} P(\eta_\gamma = u)P(X_1 + \ldots + X_r = v)P(X_{r+1} > n - (u+v)).$$

We would like to apply Esseen's theorem in order to prove that

$$P(X_1 + \ldots + X_r = v) = \frac{1}{\sigma\sqrt{r}}\,\phi\Big(\frac{v - rm}{\sigma\sqrt{r}}\Big) + O\Big(\frac{1}{r}\Big),$$

where

$$m = m(\beta) = MX = \sum t\Pi_\beta(t), \qquad \sigma^2 = \sigma^2(\beta) = M(X-m)^2. \tag{8.3}$$

Let $\beta = b_1 \ldots b_s$ and $c \neq b_s$, and consider the sequence

$$b_1 \ldots b_s \underbrace{c \ldots c}_{E \text{ times}} b_1 \ldots b_s,$$

E being a large number.

Let T_E denote the length of the shortest prefix ending with β in the word $b_2b_3 \ldots b_s c \ldots c b_1 b_2 \ldots b_s$. Since the last digit of β is different from c, we have $T_E \geq E + s$, and thus

$$T_{E+1} = T_E + 1.$$

It follows that

$$\Pi_\beta(T_E) \neq 0 \quad \text{and} \quad \Pi_\beta(T_E + 1) \neq 0, \quad \text{for every large number } E.$$

This condition guarantees as well that $\sigma \neq 0$. Note that the finiteness of the third moment is satisfied; moreover, it is true that $M(e^{\lambda X}) < \infty$ holds for a suitable positive λ. Indeed, $P(X > t)$ is the probability of the event that the sequence $\boldsymbol{\xi} = \xi_1 \dots \xi_t$ does not contain β. If it occurs, then none of $\xi_{(u-1)s+1} \dots \xi_{us}$, with $u = 1, \dots, [t/s]$, equals β, these sequences being independent; thus

$$P(X > t) \le (1 - P(\xi_1 \dots \xi_s = \beta))^{[t/s]},$$

and the assertion follows immediately.

We have thus obtained that

$$\begin{aligned}(8.4)\quad & \Pi_\gamma(r,n) \\ &= \sum_{\substack{0\le u \le c\log n\\ 0\le t\le c\log n}} P(\eta_\gamma = u)P(X>t)P(X_1+\ldots+X_r = n-u-t) + O\Big(\frac{1}{n^2}\Big) \\ &= \frac{1}{\sigma\sqrt{r}} \sum_{\substack{0\le u \le c\log n\\ 0\le t\le c\log n}} \phi\Big(\frac{(n-u-t)-mr}{\sigma\sqrt{r}}\Big)P(\eta_\gamma = u)P(X>t) \\ &\quad + O\Big(\frac{1}{r}\Big) \sum_{\substack{0\le u \le c\log n\\ 0\le t\le c\log n}} P(\eta_\gamma = u)P(X>t) + O\Big(\frac{1}{n^2}\Big).\end{aligned}$$

Clearly the last sum is $O(1)$. On the other hand, since $\phi(y_1) - \phi(y_2) = (y_1 - y_2)\phi'(y^*)$ for some $y^* \in (y_2, y_1)$, and since $\phi'(y^*) = -y^*\phi(y^*)$, it follows that

$$\left|\phi\Big(\frac{n-mr}{\sigma\sqrt{r}}\Big) - \phi\Big(\frac{n-t-mr}{\sigma\sqrt{r}}\Big)\right| \le \frac{t}{\sigma\sqrt{r}}\phi(y^*),$$

where y^* is a suitable number located between $\frac{n-t-mr}{\sigma\sqrt{r}}$ and $\frac{n-mr}{\sigma\sqrt{r}}$. This implies that the first term in (8.4) above can be written as

$$\begin{aligned}(8.5)\quad & \frac{1}{\sigma\sqrt{r}}\phi\Big(\frac{n-mr}{\sigma\sqrt{r}}\Big) \sum_{\substack{0\le u \le c\log n\\ 0\le t\le c\log n}} P(\eta_\gamma = u)P(X>t) \\ & +O\bigg(\frac{1}{\sqrt{r}}\sum_{u,t}\frac{u+t}{\sqrt{r}}\Big\{\Big|\frac{n-mr}{\sigma\sqrt{r}}\Big| + \Big|\frac{u+t}{\sigma\sqrt{r}}\Big|\Big\}\phi\Big(\frac{n-mr}{\sigma\sqrt{r}}\Big)P(\eta_\gamma{=}u)P(X>t)\bigg).\end{aligned}$$

Now observe that

$$\begin{aligned}&\sum_{\substack{0\le u \le c\log n\\ 0\le t\le c\log n}} P(\eta_\gamma = u)P(X>t) \\ &\qquad = \Big(\sum_{u=1}^{\infty} P(\eta_\gamma = u)\Big)\Big(\sum_{t=0}^{\infty} P(X>t)\Big) + O\Big(\frac{1}{n^2}\Big)\end{aligned}$$

and that

$$\sum_{u=1}^{\infty} P(\eta_\gamma = u) = P(\xi_1 \dots \xi_{s-1} = \gamma), \quad \text{while} \quad \sum_{t=0}^{\infty} P(X>t) = m.$$

Finally, the $O(\dots)$ in (8.5) is $O(1/r)$, because $|\frac{n-mr}{\sigma\sqrt{r}}|\phi(\frac{n-mr}{\sigma\sqrt{r}})$ is bounded and $\sum(u+t)^2P(\eta_\gamma = u)P(X>t) < \infty$.

From these estimates it follows that (8.4) becomes

$$\Pi_\gamma(r,n) = \frac{m}{\sigma\sqrt{r}}\phi\Big(\frac{n-mr}{\sigma\sqrt{r}}\Big) + O\Big(\frac{1}{r}\Big) + O\Big(\frac{1}{n^2}\Big).$$

We have thus proven the following result.

Lemma 5. *Assume that ξ_ν are identically distributed independent random variables, distributed as $P(\xi_\nu = j) = \delta_j$ $(j = 1, \dots, d)$, $\sum_{j=1}^d \delta_j = 1$, $\delta_j > 0$. Let β be a fixed element of $\mathcal{A}_s$, and let $\gamma \in \mathcal{A}_{s-1}$. Let m and σ be as in* (8.3)*. Denote by $\Pi_\gamma(r,n)$ the probability of the event that the random sequence $\xi_1 \dots \xi_n$ satisfies $\xi_1 \dots \xi_{s-1} = \gamma$ and that β occurs exactly r times as subword of γ. Then*

$$\Pi_\gamma(r,n) = \frac{m}{\sigma\sqrt{r}}\,\phi\Big(\frac{n-mr}{\sigma\sqrt{r}}\Big) + O\Big(\frac{1}{r}\Big) + O\Big(\frac{1}{n^2}\Big).$$

9. Proof of Theorem 5.

Every integer $n \le x$ satisfying $\omega(n) = r$ and $u_\beta(n) = l$ can be written uniquely as $n = Am$ $(\le x)$, where $P(A) \le w$ and $p(m) > w$. Now consider all the possible $\alpha \in \mathcal{A}_r$ for which $u_\beta(\alpha) = l$. The integers n satisfying $H(n) = \alpha$ are subdivided according to their corresponding number A. It is clear that A occurs in the structure of α if $H(A) \in J_\alpha$, where J_α was defined in Section 6. Let γ_A be defined by $\alpha = H(A)\gamma_A$. Further, define $J(A, \gamma_A)$ to be the occurrence of β in the sequence composed from the last $s-1$ digits of $H(A)$ concatenating with the first $s-1$ digits of γ_A. It is clear that

$$u_\beta(\alpha) = u_\beta(H(A)) + u_\beta(\gamma_A) + J(A, \gamma_A).$$

For each $\theta, \eta \in \mathcal{A}_{s-1}$, define $\mathcal{E}_\theta$ to be the set of words ending with θ, and $\mathcal{F}_\eta$ to be the set of words starting with η in the following sense. Let $\theta = e_1 \dots e_{s-1}$ and $\eta = f_1 \dots f_{s-1}$ be arbitrary elements of $\mathcal{A}_{s-1}$. We define $\mathcal{E}_\theta$ to be the set of words $\Lambda, e_{s-1}, e_{s-2}e_{s-1}, \dots, e_2 \dots e_{s-1}$ and also all γ which can be factorized as $\gamma = \tau\theta$ for some τ. On the other hand, we define $\mathcal{F}_\eta$ to be the set of words $\Lambda, f_1, f_1f_2, \dots, f_1 \dots f_{s-2}$ and also all γ which can be factorized as $\gamma = \eta\mu$ for some μ.

With this notation, we clearly have

$$M(x,r,l)=\sum_{\theta,\eta}\sum_{\substack{A\\ H(A)\in\mathcal{E}_\theta\\ P(A)\le w}}\sum_{\substack{\gamma\in\mathcal{F}_\eta\\ \lambda(\gamma)=r-\omega(A)\\ u_\beta(\gamma)=l-u_\beta(H(A))-u_\beta(\theta\eta)}} N_{r-\omega(A)}\left(\frac{x}{A}\middle|\omega;\gamma\right).$$

Assume that $r-x_2=O(x_2/x_3)$. Then, by Theorem 2,

$$M(x,r,l)=(1+o_w(1))\prod_{p\le w}\left(1-\frac{1}{p}\right)$$

$$\times\sum_{\theta,\eta}\sum_{\substack{A\\ H(A)\in\mathcal{E}_\theta\\ P(A)\le w}}\frac{1}{A}\sum_{\substack{\gamma\in\mathcal{F}_\eta\\ \lambda(\gamma)=r-\omega(A)\\ u_\beta(\gamma)=l-u_\beta(H(A))-u_\beta(\theta\eta)}}\frac{x}{x_1}t_{r-\omega(A)}(x)\varrho(\gamma).$$

Note that here we have dropped the terms corresponding to $A\gg x_2$, since their contribution was small. In this range for r, we have

$$t_{r-\omega(A)}(x)=(1+o(1))t_r(x),\qquad \text{assuming that } w=O(x_3).$$

Thus we deduce that

$$M(x,r,l)=(1+o_w(1))E_rF_l,$$

where

$$E_r:=\prod_{p\le w}\left(1-\frac{1}{p}\right)\frac{x}{x_1}t_r(x),$$

and

$$F_l:=\sum_{\theta,\eta}\sum_{\substack{H(A)\in\mathcal{E}_\theta\\ P(A)\le w}}\frac{1}{A}\Pi_\eta(l-u_\beta(H(A))-u_\beta(\theta\eta),r-\omega(A)).$$

Thus, by Lemma 5, observing that the Π_η's occurring in the sum are $(1+o(1))\Pi_\eta(l,r)$, summing on η and afterwards on A, we obtain

$$F_l=\left(\sum_{P(A)\le w}\frac{1}{A}\right)\frac{m}{\sigma\sqrt{l}}\phi\left(\frac{r-ml}{\sigma\sqrt{l}}\right)+O\left(\frac{\log w}{l}\right).$$

Since

$$\sum_{P(A)\le w}\frac{1}{A}=\prod_{p\le w}\left(1-\frac{1}{p}\right)^{-1},$$

we obtain

$$M(x,r,l)=(1+o_w(1))\frac{x}{x_1}t_r(x)\frac{m}{\sigma\sqrt{l}}\phi\left(\frac{r-ml}{\sigma\sqrt{l}}\right),$$

an estimate which is valid as $x\to\infty$, $r-x_2=O(x_2/x_3)$, $|l-x_2/m|\le x_2/x_3$.

The second formula of Theorem 5 is an easy consequence of the first one, and the third one follows immediately.

This ends the proof of Theorem 5.

10. Proof of Theorem 6. Set $w_x=x_3$, $z_x=x^{1/\sqrt{x_2}}$ and

$$E(n):=\prod_{\substack{p^\alpha\|n\\ w_x<p^\alpha<z_x}}p^\alpha.$$

Let $\mathcal{E}$ be the set of integers all prime factors of which belong to the interval (w_x,z_x).

Let $E_0,E_1,\dots,E_h\in\mathcal{E}$ be mutually coprime integers, chosen so that

$$\max_{0\le\nu\le h}\frac{\log E_\nu}{\log x}=o_x(1).$$

Then, by the sieve method, we have

$$\frac{1}{x}\#\{n\le x: E(n+j)=E_j\ (j=0,1,\dots,h)\}$$

$$=(1+o(1))\prod_{w_x<p^\alpha<z_x}\left(1-\frac{h+1}{p}\right)=(1+o(1))\prod_{w_x<p^\alpha<z_x}\left(1-\frac{1}{p}\right)^{h+1}.$$

Now let $\omega_1(n)$ be the number of prime divisors of $n/E(n)$. Since

$$\sum_{p<w_x}\frac{1}{p}+\sum_{z_x<p<x}\frac{1}{p}=O(x_3),$$

by using the Turán–Kubilius inequality, we have $\omega_1(n)=O(x_3)$ for almost all integers n. Furthermore, observe that $u_\beta(n)-u_\beta(E(n))=O(\omega_1(n))$; thus, neglecting a set of integers n having zero density, we have

$$\tau_{\beta_j}(n+j)-\tau_{\beta_j}(E(n+j))=o(1).$$

Because of this, and since Φ is continuous, it is sufficient to prove the theorem for $\tau_{\beta_j}(E(n+j))$ instead of $\tau_{\beta_j}(n+j)$.

Now it is clear that the set of integers $n\le x$ with $\max_{j=0,1,\dots,h}E(n+j)>x^{x_3/\sqrt{x_2}}$ is of zero density.

We have

$$\frac{1}{x}\#\{n\le x:\tau_{\beta_j}(E(n+j))<y_j\ (j=0,1,\dots,h)\}$$

$$=(1+o(1))\sum_{E_0,E_1,\dots,E_h}{}^{*}\frac{x}{E_0E_1\dots E_h}+o(1),$$

where the asterisk in the sum indicates that the sum is taken over those $E_0,E_1,\dots,E_h$ for which $\tau_{\beta_j}(E(n+j))<y_j$ $(j=0,1,\dots,h)$, $E_j\le x^{x_3/\sqrt{x_2}}$, $(E_i,E_j)=1$ for every $i\ne j$.

Now we can drop the condition of coprimality in Σ^*. Indeed, if we set $F = E_0E_1\dots E_h$, one can observe that F can be factorized as $E_0E_1\dots E_h$ in no more than $d_{h+1}(F)$ different ways. Hence if $(E_i, E_j) > 1$ for some $i \neq j$, then $p^2 \mid F$ for some $p > w_x$; the contribution of such p's is less than

$$\sum \frac{d_{h+1}(u)}{u} \sum \frac{d_{h+1}(v)}{v},$$

where v runs over the square-free integers with $p(v) > w_x$, $P(v) < z_x$, and u runs over the square-full integers with $p(u) > w_x$, $P(u) < z_x$, $u = 1$ being excluded. But clearly

$$\sum \frac{d_{h+1}(v)}{v} \ll \prod_{w_x<p<z_x} \left(1 + \frac{h+1}{p}\right) \ll \left(\frac{\log z_x}{\log w_x}\right)^{h+1}$$

and

$$\sum \frac{d_{h+1}(u)}{u} = \prod_{w_x<p<z_x} \left(1 + \frac{d_{h+1}(p^2)}{p^2} + \frac{d_{h+1}(p^3)}{p^3} + \dots\right) - 1$$

$$< \exp\left\{h(h+1) \sum_{p>w_x} \frac{1}{p^2}\right\} - 1 \ll \frac{1}{w_x \log w_x}.$$

Hence we have

$$\sum_{E_0,E_1,\dots,E_n}^{*} \frac{x}{E_0E_1\dots E_n} = \prod_{j=0}^{h} R(y_j) + O\left(\frac{1}{w_x \log w_x}\right),$$

where $R(y_j)$ is the number of those E_j for which $E_j < x^{x_3/\sqrt{x_2}}$ and $\tau_{\beta_j}(E_j) < y_j$.

We have thus proved that the conditions $\tau_{\beta_j}(E_j) < y_j$, $j = 0, 1, \dots, h$, are independent. It is therefore enough to prove that

$$\text{(10.1)} \qquad \frac{1}{x}\#\{n \le x : \tau_{\beta_j}(E(n+j)) < y_j\} = (1 + o(1)) \left(\frac{\log z_x}{\log w_x}\right) R(y_j)$$
$$= (1 + o(1))\Phi(y_j),$$

say $A = B = C$. But we have just proved that $A = B$. Relation (7.3) of Theorem 5 implies that

$$\frac{\log z_x}{\log w_x} R(y) = (1 + o(1))\Phi(y) \qquad \text{as } x \to \infty$$

and therefore that $B = C$. This therefore ends the proof of Theorem 6.

11. Proof of Theorem 7. Let w_x, z_x, $E(n)$, $\omega_1(n)$ and $\mathcal{E}$ be as in the proof of Theorem 6. Denote by $\Pi(x|E)$ the number of primes $p \le x$ for which $E(p+1) = E$. By the Eratosthenian sieve, we have

$$\text{(11.1)} \qquad \Pi(x|E) = \sum_{\delta} \pi(x; E\delta, -1)\mu(\delta),$$

where δ runs over the divisors of

$$K := \prod_{w_x<p<z_x} p$$

and where $\pi(x; a, b)$ stands for the number of primes $p \le x$ such that $p \equiv b \pmod a$. Using (11.1), we have

$$\text{(11.2)} \qquad \left|\Pi(x|E) - \sum_{\delta|K} \frac{\mu(\delta)}{\varphi(E\delta)}\pi(x)\right| \le \sum_{\substack{\delta|K \\ \delta \le x/E}} \left|\pi(x; E\delta, -1) - \frac{\pi(x)}{\varphi(E\delta)}\right|,$$

where φ stands for the Euler function. But it is clear that

$$\text{(11.3)} \qquad S(E) := \sum_{\delta|K} \frac{\mu(\delta)}{\varphi(E\delta)} = \frac{1}{\varphi(E)} \prod_{p|E}\left(1 - \frac{1}{p}\right) \prod_{\substack{p \nmid E \\ w_x<p<z_x}} \left(1 - \frac{1}{p-1}\right)$$
$$= \frac{1}{\varphi(E)} \cdot \frac{\log w_x}{\log z_x}\left(1 + O\left(\frac{1}{w_x}\right)\right).$$

First summing up over all $E \le x^{x_3/\sqrt{x_2}} := X$, say, we have

$$\text{(11.4)} \qquad \sum_{E<X} |\Pi(x|E) - S(E)\pi(x)| \ll \sum_{\substack{u \le x \\ u \in \mathcal{E}}} \left|\pi(x; u, -1) - \frac{\pi(x)}{\varphi(u)}\right| d(u),$$

where as usual $d(u)$ stands for the divisor function.

We shall prove that the right hand side of (11.4) is $O(x/\log^c x)$ for any given positive constant c.

We split the integers $u \le x$ into three distinct classes, namely those which are $\le X$, those satisfying $X < u \le x^{1-\varepsilon}$ and finally those such that $x^{1-\varepsilon} < u \le x$; here ε is a small positive number. We name the corresponding sums Σ_1, Σ_2 and Σ_3 respectively.

First we notice that, since $\pi(x; u, -1) \ll \pi(x)/\varphi(u)$,

$$\Sigma_2 \ll \pi(x) \sum_{\substack{u \le x^{1-\varepsilon} \\ u \in \mathcal{E}}} \frac{d(u)}{\varphi(u)}$$

and that

$$\Sigma_3 \ll x \sum_{\substack{x^{1-\varepsilon} < u \le x \\ u \in \mathcal{E}}} \frac{d(u)}{u}.$$

These sums are indeed essentially small because $P(u) < z_x$. Hence we have

$$\Sigma_2,\ \Sigma_3 = O\left(\frac{x}{\log^c x}\right).$$

We now write

$$\Sigma_1 = \sum_{\substack{d(u)<\log^B x \\ u\in\mathcal{E}}} + \sum_{\substack{d(u)>\log^B x \\ u\in\mathcal{E}}} = \Sigma_{1,1} + \Sigma_{1,2},$$

say, where B is some positive constant. We first estimate $\Sigma_{1,2}$ and obtain

$$\Sigma_{1,2} \le (\log x)^{-B}\pi(x)\sum\frac{d^2(u)}{\varphi(u)} \ll \pi(x)(\log x)^{-B}\prod_{w_x<p<z_x}\left(1+\frac{4}{p}+\dots\right)$$

$$\ll \pi(x)(\log x)^{-B}\left(\frac{\log w_x}{\log z_x}\right)^4.$$

In order to estimate $\Sigma_{1,1}$, we use the Bombieri–Vinogradov theorem in a weaker form and this allows us to obtain

$$\sum_{E<X}|\Pi(x|E) - S(E)\pi(x)| = O\left(\frac{x}{\log^c x}\right).$$

From the above estimates, it also follows that

$$\sum_{E>X}\Pi(x|E) = o(1)\pi(x).$$

By using the Turán–Kubilius inequality for the shifted primes $p+1$, we find that

$$\omega_1(p+1) = \omega\left(\frac{p+1}{E(p+1)}\right) = o(1)x_2$$

for all but $o(\pi(x))$ primes $p \le x$.

Let y be given. We shall now prove that

$$\#\left\{p \le x : \left|\frac{u_\beta(E(p+1)) - x_2/m}{c(\beta)\sqrt{x_2}}\right| < y\right\} = (1+o_x(1))\Phi(y)\pi(x).$$

In order to count the number of primes satisfying the condition contained in $\{\dots\}$, we observe that

$$(11.5)\qquad \sum_{\tau_\beta(E)<y}\Pi(x|E) = \sum_{\substack{\tau_\beta(E)<y \\ E<X}}\Pi(x|E) + o(\pi(x))$$

$$= \pi(x)\sum_{\substack{\tau_\beta(E)<y \\ E<X}}S(E) + o(\pi(x))$$

$$= \pi(x)\frac{\log w_x}{\log z_x}\sum_{\substack{\tau_\beta(E)<y \\ E<X}}\frac{1}{\varphi(E)} + o(\pi(x)),$$

where we made use of (11.3). We have thus reduced the problem to that of estimating

$$\Sigma := \sum_{\substack{\tau_\beta(E)<y \\ E<X}}\frac{1}{\varphi(E)}.$$

To do this, we use formula (7.3) of Theorem 5. First we observe that, since

$$\frac{\varphi(E)}{E} = \prod_{p|E}\left(1-\frac{1}{p}\right)$$

and writing

$$\Sigma^* := \sum_{\substack{\tau_\beta(E)<y \\ E<X}}\frac{1}{E},$$

we have

$$0 \le \Sigma - \Sigma^* = \sum_E\left(\frac{1}{\varphi(E)} - \frac{1}{E}\right) = \sum_E\frac{1}{E}\left(\prod_{p|E}\frac{1}{1-1/p} - 1\right)$$

$$= \prod_{w_x<q<z_x}\left(1+\frac{q}{(q-1)^2}\right) - \prod_{w_x<q<z_x}\frac{1}{1-1/q}$$

$$= \prod_{w_x<q<z_x}\left(1-\frac{1}{q}\right)^{-1}\left\{\prod_{w_x<q<z_x}\left(1+\frac{q}{(q-1)^2}\right)\left(1-\frac{1}{q}\right) - 1\right\}$$

$$= \left(\frac{\log z_x}{\log w_x}\right)\left\{\exp\left(\sum_{q>w_x}\frac{c}{q^2}\right) - 1\right\} \ll \left(\frac{\log z_x}{\log w_x}\right)\frac{1}{w_x}.$$

Thus

$$\sum_{\tau_\beta(E)<y}\Pi(x|E) = \pi(x)\frac{\log w_x}{\log z_x}R(y) + o(\pi(x)).$$

Hence, combining this with (11.5) and taking into account (10.1), the result follows immediately.

12. A higher dimensional problem. Let $\beta_1, \dots, \beta_h$ be given distinct words of length s and let

$$v(n) := (u_{\beta_1}(H(n)), \dots, u_{\beta_h}(H(n))).$$

By a more sophisticated method, we are able to give an asymptotic estimate for the size of integers n for which $\omega(n) = r$ and $v(n) = (l_1, \dots, l_h)$ are satisfied at least in the range $r \sim x_2$ and $l_i \sim x_2/m_i$ ($i = 1, \dots, h$). In this case, the components of $v(n)$ are dependent and their correlation depends on the special choice of β_j.

It is much easier to prove the global distribution theorem by using Lemma 3 and Theorem 2.

Let w_x be a function tending to ∞ very slowly and divide the integers $n \le x$ into classes $n = A\nu$, where $\omega(\nu) = r$, $P(A) \le w_x$, $p(\nu) > w_x$, and according to $H(\nu) = \alpha$.

For a fixed A, each α occurs $(1 + o(1))\frac{x}{x_1}t_r(x)\varrho(\alpha)$ times.

Furthermore,

$$v(n) - V(\alpha) = (O(\omega(A)), O(\omega(A)), \ldots, O(\omega(A))),$$

where

$$V(\alpha) := (u_{\beta_1}(\alpha), \ldots, u_{\beta_h}(\alpha)).$$

Consider now the random variables ξ_i which were defined in Lemma 5 and set

$$f_1(\xi_1, \ldots, \xi_s) := \begin{cases} t_\nu & \text{if } \xi_1 \ldots \xi_s = \beta_\nu \\ & \quad (\nu = 1, \ldots, h), \\ 0 & \text{if } \xi_1 \ldots \xi_s \ne \beta_\nu. \end{cases}$$

Thus

$$Mf_1(\xi_1, \ldots, \xi_s) = t_1\varrho(\beta_1) + \ldots + t_h\varrho(\beta_h) := t.$$

Then further set

$$X_j := f_1(\xi_j, \xi_{j+1}, \ldots, \xi_{j+s-1}) - t.$$

Let $\sigma = \sigma(t_1, \ldots, t_h)$ be defined by

$$\sigma^2 = MX_1^2 + 2\sum_{j=2}^{s-1} MX_1X_j.$$

It remains to prove that the quadratic form σ is positive definite.

To prove that $\sigma(t_1, \ldots, t_h) > 0$, we proceed as follows. First recall that $\xi_1, \xi_2, \ldots$ are identically distributed random variables with $P(\xi_i = l) = \delta_l$ for $l = 1, \ldots, d$. Now let f be defined on $\mathcal{A}_s$ by

$$f(\gamma) := \begin{cases} t_\nu - t & \text{if } \gamma = \beta_\nu, \\ -t & \text{otherwise.} \end{cases}$$

Let

$$Y_r = \xi_r\xi_{r+1}\ldots\xi_{r+s-1}, \qquad Z_n = f(Y_1) + \ldots + f(Y_{n-s+1}).$$

Then choose a particular $\gamma \in \mathcal{A}_s$ and consider those sequences $Y_1, \ldots, Y_{n-s+1}$ for which $Y_i = \gamma$. Let $(1 = \tau_0 <)\tau_1 < \ldots < \tau_r$ denote the sequence of the indices m for which $Y_m = \gamma$, and let

$$S_l = f(Y_{\tau_l}) + \ldots + f(Y_{\tau_{l+1}-1}) \quad (l = 0, 1, \ldots, r-1),$$
$$T = f(Y_{\tau_r}) + \ldots + f(Y_{n-s+1}).$$

Thus $Z_n = S_0 + S_1 + \ldots + S_{r-1} + T$, and the summands are mutually independent. Furthermore, all the moments of S_ν and T are finite. By using the same argument as in the proof of Lemma 5, we first find that $P(\tau_\nu = k) > 0$ for every large k. Thus $M(S_\nu - MS_\nu)^2 = \sigma_\gamma^2 > 0$ can be deduced immediately. From classical theorems on the distribution of sums of random variables with random number of summands, one can deduce that

$$\min_a M(Z_n - a)^2 > cn \quad (c > 0),$$

and hence that $\sigma(t_1, \ldots, t_h)$ is positive definite.

To prove that the limit distribution of

$$\sum_{l=1}^{h} \tau_l u_{\beta_l}(H(n)) \tag{12.1}$$

exists, and that it is the normal law with variance $\sigma(t_1, \ldots, t_h)$, we can repeat the argument used in the proof of Theorem 6.

Since the limit distribution of $v(n)$ is completely characterized by that of the projections (12.1) (see Galambos [4], Theorem 19), Theorem 8 follows immediately:

THEOREM 8. *Let*

$$F_x(y_1, \ldots, y_h) := \frac{1}{x}\#\{n \le x : \tau_{\beta_j}(n) < y_j \ (j = 1, \ldots, h)\}.$$

Then

$$\lim_{x\to\infty} F_x(y_1, \ldots, y_h) = \Phi_\sigma(y_1, \ldots, y_h),$$

where Φ_σ denotes the Gaussian law with covariance matrix corresponding to σ.

13. Additional remarks. For each $\alpha \in \mathcal{A}^*$ let $\kappa(\alpha)$ denote the largest integer k such that all possible words of length k occur as subwords in α.

To prove a sharp theorem for the order of $\kappa(H(n))$ seems to be hard in the general case. However, assuming that $\delta_1 = \ldots = \delta_d = 1/d$, we can apply the following nice result of Tamás F. Móri [7]:

If ξ_ν is an infinite sequence of independent random variables with $P(\xi_\nu = j) = 1/d$ $(j = 1, \ldots, d)$, then for every $\varepsilon > 0$, the event that

$$\left[\frac{1}{\log d}\left(\log m - \log\log m - \varepsilon\frac{\log\log m}{\log m}\right)\right]$$
$$\le \kappa(\xi_1 \ldots \xi_m) \le \left[\frac{1}{\log d}\left(\log m - \log\log m + (1+\varepsilon)\frac{\log\log m}{\log m}\right)\right]$$

holds for every large enough m, is of probability 1.

As a straightforward consequence we have the following result:

If $\delta_1 = \ldots = \delta_d = 1/d$ *and if* (1.1) *holds, then for all but* $o(x)$ *of the integers* $n \le x$*, we have*

$$\kappa(H(n)) = \frac{1}{\log d}(x_2 - x_3) + O(1).$$

This comes out by observing that $\lambda(H(n)) = x_2 + O(x_2^{3/4})$ for all but $o(x)$ of the integers $n \le x$, and using Theorem 2, taking into account that x_2 is a very slowly varying function, in the sense that $\log\log(x^c) - \log\log x = O(1)$.

References

[1] P. H. Diamanda, *Some probability limit theorems with statistical applications*, Proc. Cambridge Philos. Soc. 49 (1953), 239–246.

[2] —, *The central limit theorem for m-independent variables asymptotically stationary to second order*, ibid. 50 (1954), 287–292.

[3] C. G. Esseen, *Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law*, Acta Math. 77 (1945), 1–125.

[4] J. Galambos, *Advanced Probability Theory*, Marcel Dekker, New York, Basel, 1988.

[5] I. M. Ibrakhimov and Yu. V. Linnik, *Independent and Stationary Dependent Variables*, Nauka, Moscow, 1965.

[6] J. Kubilius, *Probabilistic Methods in the Theory of Numbers*, Amer. Math. Soc. Transl. Math. Monographs 11, Providence, 1964.

[7] T. F. Móri, *More on the waiting time till each of some given patterns occurs as a run*, Canad. J. Math. 42 (1990), 915–932.

[8] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, Inst. Élie Cartan 13, Univ. de Nancy, 1990.

DÉPARTEMENT DE MATHÉMATIQUES
ET DE STATISTIQUE
UNIVERSITÉ LAVAL
CITÉ UNIVERSITAIRE
QUÉBEC, QUÉ., CANADA G1K 7P4
E-mail:JMDK@MAT.ULAVAL.CA

INSTITUTE OF MATHEMATICS
EÖTVÖS UNIVERSITY
MÚZEUM KRT. 6-8
1088 BUDAPEST, HUNGARY

Received on 20.6.1994 (2630)