ACTA ARITHMETICA
112.4 (2004)

Arithmetic properties of numbers with restricted digits

by

William D. Banks (Columbia, MO) and
Igor E. Shparlinski (Sydney)

1. Introduction. Let $g \geq 2$ be a fixed integer base, and consider the *base-g representation* of an arbitrary integer $s \geq 0$:

$$s = \sum_{j \geq 0} d_g(j,s) g^j, \qquad d_g(j,s) \in \{0,1,\ldots,g-1\}.$$

For any fixed set of digits $\mathcal{D} \subset \{0,1,\ldots,g-1\}$ with $0 \in \mathcal{D}$, we denote by $\mathcal{A}_g(n,\mathcal{D})$ the collection of integers s such that $0 \leq s < g^n$, $\gcd(s,g) = 1$, and $d_g(j,s) \in \mathcal{D}$ for $j = 0,1,\ldots,n-1$. Clearly, one has

$$\#\mathcal{A}_g(n,\mathcal{D}) = \#\mathcal{E}(\#\mathcal{D})^{n-1},$$

where $\mathcal{E} = \{d \in \mathcal{D} \mid \gcd(d,g) = 1\}$. We also assume that the elements of $\mathcal{D}$ do not share a nontrivial common divisor, that is, $\gcd(\{d \mid d \in \mathcal{D}\}) = 1$ (otherwise most of our results fail for obvious reasons).

For the special case $g = 2$, we also consider the collection $\mathcal{B}(n,k)$ of odd integers s in the range $1 \leq s < 2^n$ such that $d_2(j,s) = 1$ for precisely k values of j, where $n \geq 4$ and $2 \leq k \leq n-2$. Since $d_2(0,s) = 1$, it is easy to see that

$$\#\mathcal{B}(n,k) = \binom{n-1}{k-1}.$$

Recall that for any fixed integer $l \geq 2$, a positive integer s is said to be *l-powerfree* if it is not divisible by p^l for any prime number p; in particular, when $l = 2$ one obtains the *squarefree* integers.

Various arithmetical questions (such as properties of divisibility, distribution in arithmetic progressions, character sums, etc.) about integers whose g-ary digits are restricted in certain ways have been considered by numerous authors; see, for example, [1, 3–13, 15, 16, 18] and the references contained therein. In this paper, we obtain asymptotic formulas for the

2000 *Mathematics Subject Classification*: 11A05, 11A63, 11N37.
Key words and phrases: Euler function, sum of divisors, powerfree integers, squarefree integers, g-ary digits.

- sum of the values of the Euler function and the sum of divisors function over $\mathcal{A}_g(n,\mathcal{D})$ and $\mathcal{B}(n,k)$,
- number of elements of $\mathcal{A}_g(n,\mathcal{D})$ and $\mathcal{B}(n,k)$ that are relatively prime to a given number $t \ge 1$,
- number of pairs of relatively prime elements in $\mathcal{A}_g(n,\mathcal{D})$ and $\mathcal{B}(n,k)$,
- number of elements of $\mathcal{A}_g(n,\mathcal{D})$ and $\mathcal{B}(n,k)$ that are l-powerfree,

and our results show that these quantities are close to their expected values for a wide range in the values of $\#\mathcal{D}$, k and n.

In particular, our Theorem 3 gives a partial answer to an open problem of Erdős, Mauduit and Sárközy (see Problem 5 in [5]). Moreover, Theorem 3 can easily be extended from $N = g^n - 1$ to arbitrary N with $g^{n-1} \le N < g^n$ since both of the main ingredients of the proof are present in this generality. Indeed, Corollary 2 to Theorem 1 of [11] holds for all N in this range, although we only use the special case $N = g^n - 1$ in our proof of Lemma 1. The estimate of Lemma 2 is an upper bound, and the analogous bound in the case where $g^{n-1} \le N < g^n$ clearly holds at the cost of an extra factor of g, which is irrelevant to our proof of Theorem 3. Hence, without substantial modifications, the same techniques solve this problem completely.

Also, our Theorem 12 provides an substantial improvement of the range of k in the Theorem 3 of [16] in the case $g = 2$ (we remark that the settings are slightly different but it seems plausible that our results can be extended to cover those of Theorem 3 of [16] in full).

Our treatment of each of these questions involves some common techniques and therefore yields results that all have a similar flavor. On the other hand, significant differences in the structure of the sets $\mathcal{A}_g(n,\mathcal{D})$ and $\mathcal{B}(n,k)$ lead to results that are nontrivial for very different ranges of the involved parameters for each of the problems we consider. We remark that for some special values of g and $\mathcal{D}$, an asymptotic formula for squarefree elements of $\mathcal{A}_g(n,\mathcal{D})$ has been given in [6].

As usual, for an integer $m \ge 1$ we denote by $\varphi(m)$ the Euler function, by $\tau(m)$ and $\sigma(m)$ the number and the sum (respectively) of distinct positive integer divisors of m, and by $\mu(m)$ the Möbius function; we recall that $\mu(1) = 1$, $\mu(m) = 0$ if m is not squarefree, and $\mu(m) = (-1)^{\omega(m)}$ otherwise, where $\omega(m)$ is the number of distinct prime divisors of $m \ge 2$. We also denote by $\zeta(l)$ (for $l \ge 2$) the special value of the Riemann zeta function:

$$\zeta(l) = \sum_{m=1}^{\infty} \frac{1}{m^l} = \prod_{p\,\mathrm{prime}} \left(1 - \frac{1}{p^l}\right)^{-1}.$$

Finally, $\ln x$ denotes the natural logarithm of a real number $x > 0$, while the notation $\log x$ is reserved for the binary logarithm: $\log x = (\ln x)/(\ln 2)$.

In what follows, we occasionally make use of the following estimates for $\varphi(m)$ and $\tau(m)$:

$$\frac{m}{\varphi(m)} = O(\ln\ln(m+2)), \qquad \ln\tau(m) = O\bigg(\frac{\ln m}{\ln\ln(m+2)}\bigg); \tag{1}$$

see Theorems 5.1 and 5.2 of Chapter 1 of [17]. We also make frequent use of the *inclusion-exclusion principle* with the Möbius function; see, for example, Theorem 2.1 in Chapter 2 of [17]. In particular, one has

$$\frac{\varphi(m)}{m} = \sum_{d|m} \frac{\mu(d)}{d}; \tag{2}$$

see (5.1) in Chapter 1 of [17].

Throughout the paper, let $g \geq 2$ and $l \geq 2$ be fixed integer parameters. All constants that occur in what follows, including any implied constants in the "O" symbol, may depend on g and l but are absolute otherwise.

We use the same symbols α and β to denote the constants that appear in all of the theorems of Section 2 and Section 3, respectively (in fact, the theorems of Section 2 hold for the same value of α and β anyway). On the other hand, we number our constants in the auxiliary statements and in the proofs since we need some specific relations among them.

Acknowledgements. The authors are very grateful to Christian Mauduit for a careful reading of the manuscript and many useful comments, in particular, for pointing out to us the relevance of the results of this paper to Problem 5 in [5]. During the preparation of this paper, W. B. was supported in part by NSF grant DMS-0070628 and I. S. was supported in part by ARC grant DP0211459. Parts of this paper were written during a visit by W. B. to Macquarie University (Sydney); the support and hospitality of this institution are gratefully acknowledged.

2. Arithmetic properties of $\mathcal{A}_g(n,\mathcal{D})$. Throughout this section, we always assume that $g \geq 3$, since our main results are trivial otherwise. Indeed, when $g = 2$ the only possible choices of $\mathcal{D}$ are $\mathcal{D} = \{0\}$ or $\mathcal{D} = \{0,1\}$; in the former case, the results are trivial, while in the latter case, $\mathcal{A}_g(n,\mathcal{D})$ is the set of all odd numbers less than 2^n and the results are well known. In particular, we use the fact that $\ln g > 1$ for $g \geq 3$.

2.1. *Divisibility of numbers from $\mathcal{A}_g(n,\mathcal{D})$.* For any $q \geq 1$ with $\gcd(q,g) = 1$, let $A_g(n,\mathcal{D};q)$ be the number of integers $s \in \mathcal{A}_g(n,\mathcal{D})$ such that $s \equiv 0 \pmod q$.

Let $\widetilde{\mathcal{A}}_g(n,\mathcal{D})$ be defined in the same way as $\mathcal{A}_g(n,\mathcal{D})$ but without the condition $\gcd(s,g) = 1$; then clearly $\#\widetilde{\mathcal{A}}_g(n,\mathcal{D}) = (\#\mathcal{D})^n$. We also define $\widetilde{A}_g(n,\mathcal{D};a,q)$ to be the number of integers $s \in \widetilde{\mathcal{A}}_g(n,\mathcal{D})$ that satisfy the

congruence $s \equiv a \pmod q$. We remark that

$$A_g(n,\mathcal{D};q) = \sum_{d\in\mathcal{E}} \widetilde{A}_g(n-1,\mathcal{D};-d\bar{g},q),$$

where $\bar{g}$ is the multiplicative inverse of g modulo q. One of our principal tools is Corollary 2 to Theorem 1 of [11], which in turn is an extension of Theorem 1 of [4]. Using that result (which applies to $\#\widetilde{\mathcal{A}}_g(n,\mathcal{D})$ and $\widetilde{A}_g(n,\mathcal{D};a,q)$ rather than $\#\mathcal{A}_g(n,\mathcal{D})$ and $A_g(n,\mathcal{D};a,q)$) and partial summation, we derive the following estimate.

LEMMA 1. *Let $g \ge 2$ be fixed, and let $\mathcal{D} \subset \{0,1,\dots,g-1\}$ be a set of digits with $0 \in \mathcal{D}$ and $\gcd(\{d \mid d \in \mathcal{D}\}) = 1$. There exists a constant $\alpha_1 > 0$ such that for all $M \le \exp(\alpha_1 n)$,*

$$\sum_{\substack{1\le q\le M\\ \gcd(q,g)=1}} \left| A_g(n,\mathcal{D};q) - \frac{\#\mathcal{A}_g(n,\mathcal{D})}{q} \right| = O(\#\mathcal{A}_g(n,\mathcal{D})\exp(-\alpha_1 n^{1/2})).$$

We also need to estimate $A_g(n,\mathcal{D};q)$ for larger values of q.

LEMMA 2. *Let $g \ge 2$ be fixed, and let $\mathcal{D} \subset \{0,1,\dots,g-1\}$ be a set of digits. Then*

$$A_g(n,\mathcal{D};q) \le g^2 q^{-\vartheta} \#\mathcal{A}_g(n,\mathcal{D}),$$

where $\vartheta = (\ln \#\mathcal{D})/(\ln g)$.

Proof. We observe that if q lies in the range $g^r \le q < g^{r+1}$ and if $q \mid s$ for some $s \in \mathcal{A}_g(n,\mathcal{D})$, then the g-ary digits of s in the rightmost r positions are uniquely determined by the g-ary digits in the leftmost $n-r$ positions. Therefore,

$$A_g(n,\mathcal{D};q) \le (\#\mathcal{D})^{n-r} \le (\#\mathcal{D})^{n+1} q^{-\vartheta}.$$

Using the estimate

$$(\#\mathcal{D})^{n+1} \le g^2(\#\mathcal{D})^{n-1} \le g^2 \#\mathcal{E}(\#\mathcal{D})^{n-1} = g^2 \#\mathcal{A}_g(n,\mathcal{D}),$$

the result follows. ∎

2.2. *Average values of φ and σ on $\mathcal{A}_g(n,\mathcal{D})$*

THEOREM 3. *Let $g \ge 2$ be fixed, and let $\mathcal{D} \subset \{0,1,\dots,g-1\}$ be a set of digits with $0 \in \mathcal{D}$ and $\gcd(\{d \mid d \in \mathcal{D}\}) = 1$. Then*

$$\sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \frac{\varphi(s)}{s} = \frac{6}{\pi^2} \prod_{\substack{p\mid g\\ p\ \text{prime}}} \frac{p^2}{p^2-1} \#\mathcal{A}_g(n,\mathcal{D})(1+O(\exp(-\alpha n^{1/2}))),$$

$$\sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \frac{\sigma(s)}{s} = \frac{\pi^2}{6} \prod_{\substack{p\mid g\\ p\ \text{prime}}} \left(1-\frac{1}{p^2}\right) \#\mathcal{A}_g(n,\mathcal{D})(1+O(\exp(-\alpha n^{1/2})))$$

for some constant $\alpha > 0$.

Proof. If the set $\mathcal{E} = \{d \in \mathcal{D} \mid \gcd(d,g) = 1\}$ is empty, then the bound is trivial; hence we may assume that $\mathcal{E} \neq \emptyset$. From (2) we obtain

$$\sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \frac{\varphi(s)}{s} = \sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \sum_{q|s} \frac{\mu(q)}{q} = \sum_{\substack{1\le q<g^n\\ \gcd(q,g)=1}} \frac{\mu(q)}{q} A_g(n,\mathcal{D};q). \tag{3}$$

Using Lemma 1 for $q \le M = \exp(\alpha_1 n)$, we see that

$$\sum_{\substack{1\le q\le M\\ \gcd(q,g)=1}} \frac{\mu(q)}{q} A_g(n,\mathcal{D};q)$$
$$= \#\mathcal{A}_g(n,\mathcal{D}) \sum_{\substack{1\le q\le M\\ \gcd(q,g)=1}} \frac{\mu(q)}{q^2} + O(\#\mathcal{A}_g(n,\mathcal{D})\exp(-\alpha_1 n^{1/2})).$$

The main term can be estimated as follows:

$$\sum_{\substack{1\le q\le M\\ \gcd(q,g)=1}} \frac{\mu(q)}{q^2} = \sum_{\substack{q\ge 1\\ \gcd(q,g)=1}} \frac{\mu(q)}{q^2} + O(M^{-1}) = \prod_{\substack{\gcd(p,g)=1\\ p\,\text{prime}}} \left(1-\frac{1}{p^2}\right) + O(M^{-1})$$
$$= \prod_{p\,\text{prime}} \left(1-\frac{1}{p^2}\right) \prod_{\substack{p|g\\ p\,\text{prime}}} \frac{p^2}{p^2-1} + O(M^{-1})$$
$$= \frac{1}{\zeta(2)} \prod_{\substack{p|g\\ p\,\text{prime}}} \frac{p^2}{p^2-1} + O(M^{-1}).$$

Therefore

$$\sum_{\substack{1\le q\le M\\ \gcd(q,g)=1}} \frac{\mu(q)}{q} A_g(n,\mathcal{D};q)$$
$$= \#\mathcal{A}_g(n,\mathcal{D})\left(\frac{6}{\pi^2} \prod_{\substack{p|g\\ p\,\text{prime}}} \frac{p^2}{p^2-1} + O(\exp(-\alpha_1 n^{1/2}))\right).$$

Hence, from (3) we derive that

$$\sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \frac{\varphi(s)}{s} = \#\mathcal{A}_g(n,\mathcal{D})\left(\frac{6}{\pi^2} \prod_{\substack{p|g\\ p\,\text{prime}}} \frac{p^2}{p^2-1} + O(\exp(-\alpha_1 n^{1/2}))\right) + O(R),$$

where

$$R = \sum_{\substack{M<q<g^n\\ \gcd(q,g)=1}} \frac{A_g(n,\mathcal{D};q)}{q}.$$

Using Lemma 2, we have

$$\begin{aligned} R &= O\Big(\#\mathcal{A}_g(n,\mathcal{D}) \sum_{M<q<g^n} q^{-1-\vartheta}\Big) \\ &= O(M^{-\vartheta}\#\mathcal{A}_g(n,\mathcal{D})) = O(\#\mathcal{A}_g(n,\mathcal{D})\exp(-\alpha_1 n^{1/2}/\ln g)). \end{aligned}$$

Taking $\alpha = \alpha_1/\ln g$, the first statement of the theorem follows.

The proof of the second statement is completely identical, except that we use the formula

$$\frac{\sigma(m)}{m} = \sum_{d|m} \frac{1}{d}$$

instead of (2); thus the main term becomes

$$\begin{aligned} \sum_{\substack{1\le q\le M\\ \gcd(q,g)=1}} \frac{1}{q^2} &= \sum_{\substack{q\ge 1\\ \gcd(q,g)=1}} \frac{1}{q^2} + O(M^{-1}) = \prod_{\substack{\gcd(p,g)=1\\ p\,\text{prime}}} \left(1-\frac{1}{p^2}\right)^{-1} + O(M^{-1}) \\ &= \zeta(2) \prod_{\substack{p|g\\ p\,\text{prime}}} \left(1-\frac{1}{p^2}\right) + O(M^{-1}), \end{aligned}$$

which finishes the proof. ∎

2.3. *Coprimality of numbers in* $\mathcal{A}_g(n,\mathcal{D})$. For any $t\ge 1$ with $\gcd(t,g)=1$, let $A_g^*(n,\mathcal{D};t)$ be the number of integers $s\in\mathcal{A}_g(n,\mathcal{D})$ such that $\gcd(s,t)=1$.

THEOREM 4. *Let $g\ge 2$ be fixed, and let $\mathcal{D}\subset\{0,1,\dots,g-1\}$ be a set of digits with $0\in\mathcal{D}$ and $\gcd(\{d\,|\,d\in\mathcal{D}\})=1$. There exists a constant $\alpha>0$ such that*

$$A_g^*(n,\mathcal{D};t) = \frac{\varphi(t)}{t}\,\#\mathcal{A}_g(n,\mathcal{D})(1+O(R)),$$

where

$$R = (\exp(-\alpha n^{1/2}) + \tau(t)\exp(-\alpha n))\ln\ln(t+2).$$

Proof. From the inclusion-exclusion principle and the relation (2), it follows that

$$A_g^*(n,\mathcal{D};t) - \frac{\varphi(t)}{t}\,\#\mathcal{A}_g(n,\mathcal{D}) = \sum_{q|t}\mu(q)\left(A_g(n,\mathcal{D};q) - \frac{\#\mathcal{A}_g(n,\mathcal{D})}{q}\right).$$

By Lemma 1, there exists a constant $\alpha_1>0$ such that

$$\begin{aligned} \left|A_g^*(n,\mathcal{D};t) - \frac{\varphi(t)}{t}\,\#\mathcal{A}_g(n,\mathcal{D})\right| &\le \sum_{\substack{q|t\\ q>M}} \left|A_g(n,\mathcal{D};q) - \frac{\#\mathcal{A}_g(n,\mathcal{D})}{q}\right| \\ &\quad + O(\#\mathcal{A}_g(n,\mathcal{D})\exp(-\alpha_1 n^{1/2})), \end{aligned}$$

where $M = \exp(\alpha_1 n)$. Using Lemma 2, we also see that

$$\sum_{\substack{q|t\\ q>M}} \left(A_g(n,\mathcal{D};q) + \frac{\#\mathcal{A}_g(n,\mathcal{D})}{q} \right) \le 2g^2\tau(t)M^{-\vartheta}\#\mathcal{A}_g(n,\mathcal{D})$$
$$= O(\tau(t)\#\mathcal{A}_g(n,\mathcal{D})\exp(-\alpha_1 n/\ln g)).$$

Taking any constant α such that $0 < \alpha < \alpha_1/\ln g$ and using (1), the result follows. ∎

Let us now define

$$\mathcal{V}_g(n,\mathcal{D}) = \{(s,t) \in \mathcal{A}_g(n,\mathcal{D}) \times \mathcal{A}_g(n,\mathcal{D}) \mid \gcd(s,t) = 1\}.$$

THEOREM 5. *Let $g \ge 2$ be fixed, and let $\mathcal{D} \subset \{0,1,\ldots,g-1\}$ be a set of digits with $0 \in \mathcal{D}$ and $\gcd(\{d \mid d \in \mathcal{D}\}) = 1$. Then*

$$\#\mathcal{V}_g(n,\mathcal{D}) = \frac{6}{\pi^2} \prod_{\substack{p|g\\ p\,prime}} \frac{p^2}{p^2-1} \#\mathcal{A}_g(n,\mathcal{D})^2(1 + O(\exp(-\alpha n^{1/2})))$$

for some constant $\alpha > 0$.

Proof. We have

$$\#\mathcal{V}_g(n,\mathcal{D}) = \sum_{s\in\mathcal{A}_g(n,\mathcal{D})} A_g^*(n,\mathcal{D};s).$$

By Theorem 4, there exists a constant $\alpha_2 > 0$ such that

$$\#\mathcal{V}_g(n,\mathcal{D}) = \sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \left(\frac{\varphi(s)}{s} \#\mathcal{A}_g(n,\mathcal{D})(1 + O(R_s)) \right),$$

where

$$R_s = (\exp(-\alpha_2 n^{1/2}) + \tau(s)\exp(-\alpha_2 n))\ln\ln(s+2).$$

Since $s < g^n$, we have $\ln\ln(s+2) = O(\ln(n+1))$, and by (1) it follows that

$$\ln\tau(s) = O\left(\frac{n}{\ln(n+1)}\right).$$

Hence for any $0 < \alpha_3 < \alpha_2$, we have

$$\#\mathcal{V}_g(n,\mathcal{D}) = \#\mathcal{A}_g(n,\mathcal{D})(1 + O(\exp(-\alpha_3 n^{1/2}))) \sum_{s\in\mathcal{A}_g(n,\mathcal{D})} \frac{\varphi(s)}{s}.$$

The result now follows from Theorem 3. ∎

An alternative way to prove Theorem 5 would be via the identity

$$\#\mathcal{V}_g(n,\mathcal{D}) = \sum_{1\le q<g^n} \mu(q)A(n,\mathcal{D};q)^2,$$

with consecutive use of Lemmas 1 and 2, as in the proof of Theorem 11 below.

2.4. *Powerfree numbers in* $\mathcal{A}_g(n,\mathcal{D})$. Let us define

$$\mathcal{P}_g(n,\mathcal{D};l) = \{s \in \mathcal{A}_g(n,\mathcal{D}) \mid s \text{ is } l\text{-powerfree}\}.$$

The following result is very similar to Theorem 4 of [4] (but not completely equivalent). We present it here for the sake of completeness.

THEOREM 6. *Let* $g \geq 2$ *be fixed, and let* $\mathcal{D} \subset \{0,1,\dots,g-1\}$ *be a set of digits with* $0 \in \mathcal{D}$, $\gcd(\{d \mid d \in \mathcal{D}\}) = 1$ *and* $\#\mathcal{D} > g^{1/l}$. *Then*

$$\#\mathcal{P}_g(n,\mathcal{D};l) = \frac{1}{\zeta(l)} \prod_{\substack{p \mid g \\ p \text{ prime}}} \frac{p^l}{p^l-1} \#\mathcal{A}_g(n,\mathcal{D})(1+O(\exp(-\alpha n^{1/2})))$$

for some constant $\alpha > 0$.

Proof. If the set $\mathcal{E} = \{d \in \mathcal{D} \mid \gcd(d,g) = 1\}$ is empty, then the bound is trivial; hence we may assume that $\mathcal{E} \neq \emptyset$. From the inclusion-exclusion principle it follows that

$$\#\mathcal{P}_g(n,\mathcal{D};l) = \sum_{\substack{1 \leq m < g^{n/l} \\ \gcd(m,g)=1}} \mu(m) A_g(n,\mathcal{D};m^l). \tag{4}$$

Using the trivial estimate

$$\sum_{\substack{1 \leq m \leq M \\ \gcd(m,g)=1}} \left| A_g(n,\mathcal{D};m^l) - \frac{\#\mathcal{A}_g(n,\mathcal{D})}{m^l} \right| \leq \sum_{\substack{1 \leq q \leq K^l \\ \gcd(q,g)=1}} \left| A_g(n,\mathcal{D};q) - \frac{\#\mathcal{A}_g(n,\mathcal{D})}{q} \right|,$$

we apply Lemma 1 with $M = K^l$, where $K = \exp(\alpha_1 n/l)$, to see that

$$\sum_{\substack{1 \leq m \leq K \\ \gcd(m,g)=1}} \mu(m) A_g(n,\mathcal{D};m^l)$$
$$= \#\mathcal{A}_g(n,\mathcal{D}) \sum_{\substack{1 \leq m \leq K \\ \gcd(m,g)=1}} \frac{\mu(m)}{m^l} + O(\#\mathcal{A}_g(n,\mathcal{D}) \exp(-\alpha_1 n^{1/2})).$$

As in the proof of Theorem 3, we can estimate

$$\sum_{\substack{1 \leq m \leq K \\ \gcd(m,g)=1}} \frac{\mu(m)}{m^l} = \frac{1}{\zeta(l)} \prod_{\substack{p \mid g \\ p \text{ prime}}} \frac{p^l}{p^l-1} + O(K^{-1}).$$

Taking $\alpha_4 = \alpha_1/l$, we have

$$\sum_{\substack{1 \leq m \leq K \\ m \text{ odd}}} \mu(m) A_g(n,\mathcal{D};m^l)$$
$$= \#\mathcal{A}_g(n,\mathcal{D}) \left(\frac{1}{\zeta(l)} \prod_{\substack{p \mid g \\ p \text{ prime}}} \frac{p^l}{p^l-1} + O(\exp(-\alpha_4 n^{1/2})) \right).$$

Hence, from (4) we derive that

$$\begin{aligned}&\#\mathcal{P}_g(n,\mathcal{D};l)\\&\quad=\#\mathcal{A}_g(n,\mathcal{D})\Bigg(\frac{1}{\zeta(l)}\prod_{\substack{p\mid g\\ p\,\text{prime}}}\frac{p^l}{p^l-1}+O(\exp(-\alpha_4 n^{1/2}))\Bigg)+O(R),\end{aligned}$$

where

$$R=\sum_{\substack{K<m<g^{n/l}\\ \gcd(m,g)=1}}A_g(n,\mathcal{D};m^l).$$

The condition $\#\mathcal{D}>g^{1/l}$ guarantees that $l\vartheta>1$. Using Lemma 2, we can estimate the error term as follows:

$$R=O\Bigg(\#\mathcal{A}_g(n,\mathcal{D})\sum_{K<m<g^{n/l}}m^{-l\vartheta}\Bigg)=O(K^{1-l\vartheta}\#\mathcal{A}_g(n,\mathcal{D})).$$

Thus, with the choice

$$\alpha=\min\left\{\alpha_4,\frac{\alpha_1(l\vartheta-1)}{l}\right\},$$

we obtain the stated result. ∎

For example, for the base $g=3$ and the set of digits $\mathcal{D}=\{0,1\}$, we find that the density of squarefree numbers in $\mathcal{A}_3(n,\{0,1\})$ is approximately $27/(4\pi^2)\approx 0.6839$ as $n\to\infty$.

3. Arithmetic properties of $\mathcal{B}(n,k)$. Throughout this section, for simplicity, we formulate and prove our results only for the case $k\le n/2$. However, since the sets $\mathcal{B}(n,k)$ and $\mathcal{B}(n,n-k)$ are virtually indistinguishable from an arithmetic standpoint, the case where $k\ge n/2$ can be handled simply by replacing k with $n-k$ everywhere in the statements of our results and in their proofs, as the reader may easily verify.

3.1. *Divisibility of numbers from $\mathcal{B}(n,k)$.* For any $q\ge 1$, let $B(n,k;q)$ be the number of integers $s\in\mathcal{B}(n,k)$ such that $s\equiv 0 \pmod q$.

The following statement follows from Theorem 2 of [16].

LEMMA 7. *There exists an absolute constant $\beta_1>0$ such that*

$$\left|B(n,k;q)-\frac{\#\mathcal{B}(n,k)}{q}\right|=O\Bigg(\frac{\#\mathcal{B}(n,k)}{q}\exp(-\beta_1 k/\ln q)\Bigg)$$

for all $q\le M$, where $M=\exp(\beta_1 k^{1/2})$.

We also need to estimate $B(n,k;q)$ for larger values of q.

LEMMA 8. *If* $q \le 2^{n-2k+2}$, *then*

$$B(n,k;q) \le ekq^{-\vartheta}\#\mathcal{B}(n,k), \quad \textit{where} \quad \vartheta = \frac{k-1}{(n-1)\ln 2}.$$

Proof. We observe that if q lies in the range $2^r \le q < 2^{r+1}$ and if $q \mid s$ for some $s \in \mathcal{B}(n,k)$, then the binary digits of s in the rightmost r positions are uniquely determined by the binary digits in the leftmost $n-r$ positions. Recalling that s is odd, it therefore follows that

$$B(n,k;q) \le \sum_{j=0}^{k-1} \binom{n-r}{j}. \tag{5}$$

Now the inequality $1-x \le \exp(-x)$ holds for any $x \ge 0$; in particular,

$$\frac{n-r-j}{n-1-j} \le \exp\left(-\frac{r-1}{n-1-j}\right) \le \exp\left(-\frac{r-1}{n-1}\right)$$

for $j = 0, 1, \ldots, k-2$, and it follows that

$$\binom{n-r}{k-1} \le \binom{n-1}{k-1} \exp\left(-\frac{(k-1)(r-1)}{n-1}\right). \tag{6}$$

Because $r \le \log q \le n-2k+2$, from (5) we obtain the estimate

$$B(n,k;q) \le k2^{-(r-1)\vartheta}\binom{n-1}{k-1} \le ekq^{-\vartheta}\binom{n-1}{k-1} = ekq^{-\vartheta}\#\mathcal{B}(n,k)$$

(here we have used the fact that $\vartheta < 1/\ln 4$). ∎

For very large values of q (e.g., for $q > 2^{n-2k+2}$), the trivial estimate

$$B(n,k;q) < 2^n/q \tag{7}$$

suffices for our applications.

3.2. *Average values of* φ *and* σ *on* $\mathcal{B}(n,k)$

THEOREM 9. *For some constant* $\beta > 0$,

$$\sum_{s\in\mathcal{B}(n,k)} \frac{\varphi(s)}{s} = \frac{8}{\pi^2}\#\mathcal{B}(n,k)(1+O(\exp(-\beta k^{3/2}n^{-1}))),$$

$$\sum_{s\in\mathcal{B}(n,k)} \frac{\sigma(s)}{s} = \frac{\pi^2}{8}\#\mathcal{B}(n,k)(1+O(\exp(-\beta k^{3/2}n^{-1})))$$

for some constant $\alpha > 0$.

Proof. From (2) we obtain

$$\sum_{s\in\mathcal{B}(n,k)} \frac{\varphi(s)}{s} = \sum_{s\in\mathcal{B}(n,k)} \sum_{q\mid s} \frac{\mu(q)}{q} = \sum_{\substack{1\le q<2^n \\ q\,\mathrm{odd}}} \frac{\mu(q)}{q} B(n,k;q). \tag{8}$$

Using Lemma 7 for $q \le M = \exp(\beta_1 k^{1/2})$, we see that

$$\sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \frac{\mu(q)}{q} B(n,k;q) = \#\mathcal{B}(n,k) \sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \frac{\mu(q)}{q^2} + O(\#\mathcal{B}(n,k)\exp(-k^{1/2})).$$

As in the proof of Theorem 3, we can estimate

$$\sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \frac{\mu(q)}{q^2} = \frac{4}{3\zeta(2)} + O(M^{-1}) = \frac{8}{\pi^2} + O(\exp(-\beta_1 k^{1/2})). \tag{9}$$

Taking $\beta_2 = \min\{\beta_1, 1\}$, it follows that

$$\sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \frac{\mu(q)}{q} B(n,k;q) = \#\mathcal{B}(n,k)\left(\frac{8}{\pi^2} + O(\exp(-\beta_2 k^{1/2}))\right).$$

Hence, from (8) we derive that

$$\sum_{s\in\mathcal{B}(n,k)} \frac{\varphi(s)}{s} = \#\mathcal{B}(n,k)\left(\frac{8}{\pi^2} + O(\exp(-\beta_2 k^{1/2}))\right) + O(R),$$

where

$$R = \sum_{\substack{M<q<2^n\\ q\,\mathrm{odd}}} \frac{B(n,k;q)}{q}.$$

Using Lemma 8 and the estimate (11), we have

$$\begin{aligned} R &= O\Big(\#\mathcal{B}(n,k)k \sum_{M<q<2^n} q^{-1-\vartheta}\Big) = O(kM^{-\vartheta}\#\mathcal{B}(n,k)) \\ &= O(k\exp(-\beta_1 k^{3/2} n^{-1})\#\mathcal{B}(n,k)). \end{aligned}$$

Choosing any β such that $0 < \beta < \beta_2$, we obtain the first statement of the theorem.

As in the proof of Theorem 3, the proof of the second statement is completely identical except for the constant appearing in (9), which now becomes $3\zeta(2)/4$. ∎

3.3. *Coprimality of numbers in* $\mathcal{B}(n,k)$. For any odd integer $t \ge 1$, let $B^*(n,k;t)$ be the number of integers $s \in \mathcal{B}(n,k)$ such that $\gcd(s,t) = 1$.

THEOREM 10. *There exists a constant* $\beta > 0$ *such that*

$$B^*(n,k;t) = \frac{\varphi(t)}{t}\,\#\mathcal{B}(n,k)(1 + O(R)),$$

where

$$R = \exp(-\beta k^{3/2} n^{-1})\tau(t)\ln\ln(t+2).$$

Proof. From the inclusion-exclusion principle, we see that

$$B^*(n,k;t) = \sum_{q|t} \mu(q) B(n,k;q). \tag{10}$$

Let $\beta_1 > 0$ be selected as in Lemma 7, and put $M = \exp(\beta_1 k^{1/2})$. We consider two separate cases.

First, suppose that $M < 2^{n-2k+1}$, and put $L = 2^{n-2k+1}$. We apply Lemma 7 for $q \le M$, Lemma 8 for $M < q \le L$, and the bound (7) for $q > L$. Therefore, from (10) we derive that

$$B^*(n,k;t) = \#\mathcal{B}(n,k) \sum_{\substack{q|t \\ q\le M}} \frac{\mu(q)}{q} + O(R_1 + R_2 + R_3),$$

where

$$R_1 = \#\mathcal{B}(n,k) \sum_{\substack{q|t \\ q\le M}} \exp(-\beta_1 k/\ln q),$$

$$R_2 = \#\mathcal{B}(n,k) k \sum_{\substack{q|t \\ M<q\le L}} q^{-\vartheta}, \qquad R_3 = 2^n \sum_{\substack{q|t \\ q>L}} q^{-1}.$$

Using (2), we have

$$\sum_{\substack{q|t \\ q\le M}} \frac{\mu(q)}{q} = \frac{\varphi(t)}{t} + O(\tau(t)M^{-1}) = \frac{\varphi(t)}{t} + O(\tau(t)\exp(-\beta_1 k^{1/2})).$$

Also,

$$\frac{R_1}{\#\mathcal{B}(n,k)} = O(\tau(t)\exp(-\beta_1 k/\ln M)) = O(\tau(t)\exp(-k^{1/2})).$$

Next, for $k \ge 2$ we have $(k-1)/(n-1) \ge (k-1)/n \ge k/2n$, thus

$$M^{\vartheta} = \exp(\beta_1 \vartheta k^{1/2}) \ge \exp\left(\frac{\beta_1 k^{3/2}}{2n\ln 2}\right) \ge \exp(\beta_1 k^{3/2} n^{-1}). \tag{11}$$

Hence

$$\frac{R_2}{\#\mathcal{B}(n,k)} = O(\tau(t) k M^{-\vartheta}) = O(\tau(t) k \exp(-\beta_1 k^{3/2} n^{-1})).$$

Finally, using (6) with $r = n - 2k + 2$, we have

$$\begin{aligned} 2^{2k-2} &\le (2k-2)\binom{2k-2}{k-1} \le (2k-2)\binom{n-1}{k-1}\exp\left(-\frac{(n-2k+1)(k-1)}{n-1}\right) \\ &= (2k-2)\#\mathcal{B}(n,k)2^{-\vartheta(n-2k+1)} \le (2k-2)\#\mathcal{B}(n,k)M^{-\vartheta} \\ &\le (2k-2)\#\mathcal{B}(n,k)\exp(-\beta_1 k^{3/2} n^{-1}). \end{aligned}$$

Then

$$R_3 = O(\tau(t)2^n L^{-1}) = O(\tau(t)2^{2k}) = O(\tau(t)k\exp(-\beta_1 k^{3/2}n^{-1})\#\mathcal{B}(n,k)).$$

Choosing any β such that $0 < \beta < \min\{\beta_1, 1\}$ and using the above estimates together with (1), the result follows for the case $M < 2^{n-2k+1}$.

Next we turn to the case $M \geq 2^{n-2k+1}$. Taking $L = M$ and proceeding as before with the same choice of β, we see that

$$R_1 = O(\tau(t)\exp(-k^{1/2})\#\mathcal{B}(n,k)), \qquad R_2 = 0,$$
$$R_3 = O(\tau(t)2^n M^{-1}) = O(\tau(t)\exp(-\beta_1 k^{1/2})2^n).$$

Consequently,

$$B^*(n,k;t) = \frac{\varphi(t)}{t}\,\#\mathcal{B}(n,k) + O(\tau(t)\exp(-\beta k^{1/2})(\#\mathcal{B}(n,k) + 2^n)).$$

We claim that

$$2^n = O\left(n^{1/2}\binom{n-1}{k-1}\right) = O(n^{1/2}\#\mathcal{B}(n,k)) \tag{12}$$

in this case. Assuming that (12) is correct we see that in this case the result also holds with any β such that $0 < \beta < \min\{\beta_1, 1\}$.

To prove (12), we use the following estimate (see Lemma 8 of [14]):

$$\binom{n-1}{k-1} \geq (2n)^{-1/2}2^{(n-1)H((k-1)/(n-1))}, \tag{13}$$

where

$$H(x) = -x\log x - (1-x)\log(1-x), \qquad 0 < x < 1.$$

Note that $H(x)$ is strictly increasing for $0 < x < 1/2$, as is easily verified. From the estimate $\log(1 \pm \eta) = O(\eta)$, it follows that

$$H\left(\frac{1-\eta}{2}\right) = 1 + O(\eta^2), \qquad 0 < \eta < 1.$$

From the condition on M, we derive the inequality

$$\frac{1}{2} \geq \frac{k-1}{n-1} \geq \frac{1}{2} - \frac{\log M}{2(n-1)} = \frac{1-\eta}{2},$$

where

$$\eta = \frac{\log M}{n-1} = O(n^{-1/2}).$$

Therefore,

$$H\left(\frac{k-1}{n-1}\right) \geq H\left(\frac{1-\eta}{2}\right) = 1 + O(n^{-1}).$$

Thus, from (13) we see that (12) holds, which completes the proof. ∎

Let us now define

$$\mathcal{W}(n,k) = \{(s,t) \in \mathcal{B}(n,k) \times \mathcal{B}(n,k) \mid \gcd(s,t) = 1\}.$$

THEOREM 11. *Let $\delta > 0$ be fixed, and let k and n be integers such that*

$$\frac{\ln 2 + \delta}{2} \le \frac{k}{n} \le \frac{1}{2}.$$

Then

$$\#\mathcal{W}(n,k) = \frac{8}{\pi^2}\,\#\mathcal{B}(n,k)^2(1 + O(\exp(-\beta\delta n^{-1/2})))$$

for some constant $\beta > 0$.

Proof. We start with the identity

$$\#\mathcal{W}(n,k) = \sum_{\substack{1\le q<2^n\\ q\,\mathrm{odd}}} \mu(q)B(n,k;q)^2, \tag{14}$$

which follows from the inclusion-exclusion principle.

To estimate the main term, let β_1 be selected as in Lemma 7, and put $M = \exp(\beta_1 k^{1/2})$. Applying Lemma 7 for $q \le M$, we see that

$$\begin{aligned}\sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \mu(q)B(n,k;q)^2 &= \sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \mu(q)\left(\frac{\#\mathcal{B}(n,k)}{q}\,(1 + O(\exp(-\beta_1 k/\ln q)))\right)^2\\ &= \#\mathcal{B}(n,k)^2 \sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \left(\frac{\mu(q)}{q^2} + O(q^{-2}\exp(-k^{1/2}))\right).\end{aligned}$$

By (9), it follows that

$$\sum_{\substack{1\le q\le M\\ q\,\mathrm{odd}}} \mu(q)B(n,k;q)^2 = \frac{8}{\pi^2}\,\#\mathcal{B}(n,k)^2(1 + O(\exp(-\beta_2 k^{1/2}))),$$

where $\beta_2 = \min\{\beta_1, 1\}$ as in Theorem 9. Hence, from (14) we derive that

$$\#\mathcal{W}(n,k) = \frac{8}{\pi^2}\,\#\mathcal{B}(n,k)^2(1 + O(\exp(-\beta_2 k^{1/2}))) + O(R), \tag{15}$$

where

$$R = \sum_{\substack{M<q<2^n\\ q\,\mathrm{odd}}} B(n,k;q)^2.$$

We now turn to the estimation of the error term R. Let $L = 2^{n-2k+2}$ be fixed in what follows.

First, we consider the case $M < L$. In this case, we split R into two summations,

$$R = R_1 + R_2,$$

where

$$R_1 = \sum_{\substack{M<q\le L\\ q\,\mathrm{odd}}} B(n,k;q)^2, \qquad R_2 = \sum_{\substack{L<q\le 2^n\\ q\,\mathrm{odd}}} B(n,k;q)^2.$$

Taking into account the lower bound specified for k, we have

$$\begin{aligned} 1-2\vartheta &= 1-\frac{2(k-1)}{(n-1)\ln 2} = 1-\frac{2k}{n\ln 2}+O(n^{-1}) \\ &\le 1-\frac{\ln 2+\delta}{\ln 2}+O(n^{-1}) = -\frac{\delta}{\ln 2}+O(n^{-1}) \le -\delta \end{aligned}$$

provided that n is large enough. Using Lemma 8 for $M<q\le L$, we have

$$\begin{aligned} R_1 &= O\Big(k^2\#\mathcal{B}(n,k)^2 \sum_{M<q\le L} q^{-2\vartheta}\Big) = O(k^2M^{1-2\vartheta}\#\mathcal{B}(n,k)^2) \\ &= O(k^2M^{-\delta}\#\mathcal{B}(n,k)^2) = O(\exp(-\beta_3\delta k^{1/2})\#\mathcal{B}(n,k)^2), \end{aligned}$$

where β_3 is any constant such that $0<\beta_3<\beta_2$. Using the trivial bound (7) for $m>L$, we also have

$$R_2 = O\Big(\sum_{L<m<2^n}\frac{2^{2n}}{q^2}\Big) = O(2^{2n}L^{-1}) = O(2^{4k}L).$$

As before, we have

$$\begin{aligned} 2^{2k-2} &\le (2k-2)\binom{2k-2}{k-1} \le (2k-2)\binom{n-1}{k-1}\exp\Big(-\frac{(n-2k+1)(k-1)}{n-1}\Big) \\ &= (2k-2)2^{-\vartheta(n-2k+1)}\#\mathcal{B}(n,k) \le (2k-2)e^{1/2}L^{-\vartheta}\#\mathcal{B}(n,k). \end{aligned}$$

Therefore,

$$\begin{aligned} R_2 &= O(k^2L^{1-2\vartheta}\#\mathcal{B}(n,k)^2) = O(k^2M^{1-2\vartheta}\#\mathcal{B}(n,k)^2) \\ &= O(\exp(-\beta_3\delta k^{1/2})\#\mathcal{B}(n,k)^2). \end{aligned}$$

Thus, for $M<L$, we have

$$R = R_1+R_2 = O(\exp(-\beta_3\delta k^{1/2})\#\mathcal{B}(n,k)^2).$$

Combining this estimate with (15), the proof is completed in this case.

Next, we turn to the case $M\ge L$. Using the trivial bound (7) for $q>M$, we have

$$R = O\Big(\sum_{M<q<2^n}\frac{2^{2n}}{q^2}\Big) = O(2^{2n}M^{-1}) = O(2^{2n}\exp(-\beta_1k^{1/2})).$$

The condition $M\ge L$ is equivalent to $(n-2k+2)\ln 2\le \beta_1k^{1/2}$, which implies the inequality

$$\frac{1}{2}\ge\frac{k-1}{n-1}>\frac{1}{2}-\frac{\beta_1k^{1/2}}{2(n-1)\ln 2} = \frac{1-\eta}{2},$$

where

$$\eta = \frac{\beta_1 k^{1/2}}{(n-1)\ln 2} = O(n^{-1/2}).$$

Thus, the estimate (12) is valid in this situation, and the result follows. ∎

An alternative way to study $\#\mathcal{W}(n,k)$ would be via the identity

$$\#\mathcal{W}(n,k) = \sum_{s\in\mathcal{B}(n,k)} B^*(n,k;s),$$

making use of Theorem 10 as in the proof of Theorem 5. However, in the case of the set $\mathcal{B}(n,k)$, this approach leads to a much weaker result.

3.4. *Powerfree numbers in $\mathcal{B}(n,k)$.* Let us define

$$\mathcal{Q}(n,k;l) = \{s \in \mathcal{B}(n,k) \mid s \text{ is } l\text{-powerfree}\}.$$

THEOREM 12. *Let $l \ge 2$ and $\delta > 0$ be fixed, and let k and n be integers such that*

$$\frac{\ln 2 + \delta}{l} \le \frac{k}{n} \le \frac{1}{2}.$$

Then

$$\#\mathcal{Q}(n,k;l) = \frac{1}{\zeta(l)}\,\frac{2^l}{2^l-1}\,\#\mathcal{B}(n,k)(1 + O(\exp(-\beta\delta n^{1/2})))$$

for some constant $\beta > 0$.

Proof. As before, let $\beta_2 = \min\{\beta_1, 1\}$ where β_1 is selected as in Lemma 7. Put $M = \exp(\beta_2 k^{1/2}/l)$. As in the proof of Theorem 6, we begin with a relation of the form

$$\#\mathcal{Q}(n,k;l) = \sum_{\substack{1\le m<2^{n/l}\\ m\,\text{odd}}} \mu(m) B(n,k;m^l). \tag{16}$$

Applying Lemma 7 for $m \le M$, we see that

$$\sum_{\substack{1\le m\le M\\ m\,\text{odd}}} \mu(m)B(n,k;m^l) = \#\mathcal{B}(n,k) \sum_{\substack{1\le m\le M\\ m\,\text{odd}}} \left(\frac{\mu(m)}{m^l} + O(\exp(-\beta_1 k/\ln m^l))\right).$$

As in the proof of Theorem 3, we have the estimate

$$\sum_{\substack{1\le m\le M\\ m\,\text{odd}}} \frac{\mu(m)}{m^l} = \frac{1}{\zeta(l)}\,\frac{2^l}{2^l-1} + O(M^{-1}) = \frac{1}{\zeta(l)}\,\frac{2^l}{2^l-1} + O(\exp(-\beta_2 k^{1/2}/l)).$$

We also have

$$\sum_{\substack{1\le m\le M\\ m\,\mathrm{odd}}} \exp(-\beta_1 k/\ln m^l) \le M\exp(-\beta_1 k/\ln M^l)$$
$$= O(\exp((\beta_2/l-\beta_1/\beta_2)k^{1/2})).$$

For our choice of β_2, we have $2\beta_2/l \le \beta_2 \le \beta_1/\beta_2$, thus

$$0<\beta_2/l\le \beta_1/\beta_2-\beta_2/l.$$

Consequently,

$$\sum_{\substack{1\le m\le M\\ m\,\mathrm{odd}}} \mu(m)B(n,k;m^l) = \frac{1}{\zeta(l)}\frac{2^l}{2^l-1}\#\mathcal{B}(n,k)(1+O(\exp(-\beta_4 k^{1/2}))),$$

where β_4 is any constant such that $0<\beta_4<\beta_2/l$. Hence, from (16) we derive that

$$\#\mathcal{Q}(n,k;l) = \frac{1}{\zeta(l)}\frac{2^l}{2^l-1}\#\mathcal{B}(n,k)(1+O(\exp(-\beta_4 k^{1/2}))) + O(R), \tag{17}$$

where

$$R = \sum_{\substack{M<m<2^{n/l}\\ m\,\mathrm{odd}}} B(n,k;m^l).$$

We now turn to the estimation of the error term R. Let $L=2^{(n-2k+2)/l}$ be fixed in what follows.

First, we consider the case $M<L$. In this case, we split R into two summations,

$$R = R_1+R_2,$$

where

$$R_1 = \sum_{\substack{M<m\le L\\ m\,\mathrm{odd}}} B(n,k;m^l), \qquad R_2 = \sum_{\substack{L<m\le 2^{n/l}\\ m\,\mathrm{odd}}} B(n,k;m^l).$$

Taking into account the lower bound specified for k, one has

$$1-l\vartheta = 1-\frac{(k-1)l}{(n-1)\ln 2} = 1-\frac{kl}{n\ln 2}+O(n^{-1})$$
$$\le 1-\frac{\ln 2+\delta}{\ln 2}+O(n^{-1}) = -\frac{\delta}{\ln 2}+O(n^{-1}) \le -\delta$$

provided that n is large enough. Using Lemma 8 for $M<m\le L$, we have

$$R_1 = O\Big(k\#\mathcal{B}(n,k)\sum_{M<m\le L} m^{-l\vartheta}\Big) = O(kM^{1-l\vartheta}\#\mathcal{B}(n,k))$$
$$= O(kM^{-\delta}\#\mathcal{B}(n,k)) = O(\exp(-\beta_4\delta k^{1/2})\#\mathcal{B}(n,k)).$$

Using the trivial bound (7) for $m > L$, we also have

$$R_2 = O\Bigg(\sum_{L<m<2^{n/l}} \frac{2^n}{m^l}\Bigg) = O(2^n L^{1-l}) = O(2^{2k} L).$$

As before, we have

$$\begin{aligned} 2^{2k-2} &\le (2k-2)\binom{2k-2}{k-1} \le (2k-2)\binom{n-1}{k-1}\exp\Bigg(-\frac{(n-2k+1)(k-1)}{n-1}\Bigg) \\ &= (2k-2)2^{-\vartheta(n-2k+1)}\#\mathcal{B}(n,k) \le (2k-2)e^{1/2}L^{-l\vartheta}\#\mathcal{B}(n,k). \end{aligned}$$

Therefore,

$$\begin{aligned} R_2 &= O(kL^{1-l\vartheta}\#\mathcal{B}(n,k)) = O(kM^{1-l\vartheta}\#\mathcal{B}(n,k)) \\ &= O(\exp(-\beta_4\delta k^{1/2})\#\mathcal{B}(n,k)). \end{aligned}$$

Thus, for $M < L$, we have

$$R = R_1 + R_2 = O(\exp(-\beta_4\delta k^{1/2})\#\mathcal{B}(n,k)).$$

Combining this estimate with (17), the proof is completed in this case.

Next, we turn to the case $M \ge L$. Using the trivial bound (7) for $m > M$, we have

$$R = O\Bigg(\sum_{M<m<2^{n/l}} \frac{2^n}{m^l}\Bigg) = O(2^n M^{1-l}) = O(2^n M^{-1}) = O(2^n \exp(-\beta_1 k^{1/2})).$$

The condition $M \ge L$ is equivalent to $(n-2k+2)\ln 2 \le \beta_2 k^{1/2}/l$, which implies the inequality

$$\frac{1}{2} \ge \frac{k-1}{n-1} > \frac{1}{2} - \frac{\beta_2 k^{1/2}}{2(n-1)l\ln 2} = \frac{1-\eta}{2},$$

where

$$\eta = \frac{\beta_2 k^{1/2}}{(n-1)l\ln 2} = O(n^{-1/2}).$$

Thus, the estimate (12) is valid in this situation, and the result follows. ∎

In view of the comments given at the beginning of this section, we now see, for example, that in the special case $l = 2$, the density of squarefree numbers in $\mathcal{B}(n,k)$ is close to $8/\pi^2 \approx 0.8106$ for $0.35 < k/n \le 0.65$ provided that n is sufficiently large.

4. Remarks. Analogues of Theorems 3 and 9 also hold for $\sigma(s)/s$ where $\sigma(s)$ is the sum of divisors function. However, we do not have any nontrivial bounds for sums of $\varphi(s)$, $\sigma(s)$ and $\tau(s)$ over the elements of the sets $\mathcal{A}_g(n;\mathcal{D})$ and $\mathcal{B}(n,k)$.

Unfortunately, it is probably infeasible nowadays to study prime values in $\mathcal{A}_g(n;\mathcal{D})$ and $\mathcal{B}(n,k)$. However, it should be possible to study some

other interesting subsets of $\mathcal{A}_g(n;\mathcal{D})$ and $\mathcal{B}(n,k)$ that are defined by certain arithmetic properties.

It would be very interesting to determine the analogue of our results for smaller (or larger) values of k. Results in the case $k = o(n)$ would be particularly useful and interesting.

One might try to address similar questions for integers whose digits in a certain base are restricted in other ways.

For example, it would be interesting to extend the results of Section 3 to collections of integers with a fixed sum of digits in an arbitrary base $g \geq 2$. Although the analogue of our basic tool Lemma 7 is known in this setting (see Theorem 2 of [16]), working with formulas for the cardinalities of such sets and other associated estimates (such as the analogue of Lemma 8) would be technically rather complicated.

Recently some arithmetic properties of palindromes have been studied in [2], including asymptotic formulas for palindromes in arithmetic progressions and an upper bound on the number of prime values. It is certainly very natural to try to establish analogues of our results for palindromes as well.

References

[1] W. Banks, A. Conflitti and I. E. Shparlinski, *Character sums over integers with restricted g-ary digits*, Illinois J. Math. 46 (2002), 819–836.

[2] W. Banks, D. Hart and M. Sakata, *Almost all palindromes are composite*, preprint, 2003.

[3] W. Banks and I. E. Shparlinski, *On the number of sparse RSA exponents*, J. Number Theory 95 (2002), 340–350.

[4] P. Erdős, C. Mauduit and A. Sárközy, *On arithmetic properties of integers with missing digits*, *I*, ibid. 70 (1998), 99–120.

[5] —, —, —, *On arithmetic properties of integers with missing digits*, *II*, Discrete Math. 200 (1999), 149–164.

[6] M. Filaseta and S. Konyagin, *Squarefree values of polynomials all of whose coefficients are* 0 *and* 1, Acta Arith. 74 (1996), 191–205.

[7] E. Fouvry et C. Mauduit, *Méthodes de crible et fonctions sommes des chiffres*, ibid. 77 (1996), 339–351.

[8] —,—, *Sommes des chiffres et nombres presque premiers*, Math. Ann. 305 (1996), 571–599.

[9] J. B. Friedlander and I. E. Shparlinski, *On the distribution of Diffie–Hellman triples with sparse exponents*, SIAM J. Discrete Math. 14 (2001), 162–169.

[10] A. O. Gel'fond, *Sur les nombres qui ont des propriétés additives et multiplicatives données*, Acta Arith. 13 (1968), 259–265.

[11] S. Konyagin, *Arithmetic properties of integers with missing digits*: *distribution in residue classes*, Period. Math. Hungar. 42 (2001), 145–162.

[12] S. Konyagin, C. Mauduit and A. Sárközy, *On the number of prime factors of integers characterized by digit properties*, ibid. 40 (2000), 37–52.

[13] F. Luca, *Arithmetic properties of positive integers with fixed digit sum*, preprint, 2002.
[14] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
[15] C. Mauduit and A. Sárközy, *On the arithmetic structure of sets characterized by sum of digits properties*, J. Number Theory 61 (1996), 25–38.
[16] —, —, *On the arithmetic structure of the integers whose sum of digits is fixed*, Acta Arith. 81 (1997), 145–173.
[17] K. Prachar, *Primzahlverteilung*, Springer, Berlin, 1957.
[18] I. E. Shparlinski, *Prime divisors of sparse integers*, Period. Math. Hungar. 46 (2003), 215–222.

Department of Mathematics
University of Missouri
Columbia, MO 65211, U.S.A.
E-mail: bbanks@math.missouri.edu

Department of Computing
Macquarie University
Sydney, NSW 2109, Australia
E-mail: igor@ics.mq.edu.au

Received on 20.12.2002
and in revised form on 15.9.2003 (4431)