SUBBLOCK OCCURRENCES IN SIGNED DIGIT REPRESENTATIONS

PETER J. GRABNER†, CLEMENS HEUBERGER‡, AND HELMUT PRODINGER

Abstract. Signed digit representations with base q and digits $-\frac{q}{2}, \dots, \frac{q}{2}$ (and uniqueness being enforced by applying a special rule which decides whether $-q/2$ or $q/2$ should be taken) are considered with respect to counting the occurrences of a given (contiguous) subblock of length r. The average number of occurrences amongst the numbers $0, \dots, n-1$ turns out to be $const \cdot \log_q n + \delta(\log_q n) + o(1)$, with a constant and a periodic function of period one depending on the given subblock; they are explicitly described. Furthermore, we use probabilistic techniques to prove a central limit theorem for the number of occurrences of a given subblock.

1. Introduction

If we write 10^6 in binary we obtain 11110100001001000000. This *word* contains the (contiguous) substring 100 three times. In this paper we are concerned with counting occurrences of a given substring (or block) (like 100) in representations of numbers. Since this is typically somewhat erratic, we are interested in an average

$$\frac{1}{N} \sum_{0 \le n < N} (\text{number of occurrences of a given block } w \text{ in the representation of } n).$$

This is a generalisation of counting the frequency of digits.

For the instance of the q–ary representation of numbers this average was investigated by Kirschenhofer in [11]; the more exotic (q, d)–ary representation of numbers with base q and digits $d, d+1, \dots, d+q-1$ was treated in [12]. The technique in these papers was an extension of Delange's method [3]. However, in [6] a novel method, based on the Mellin–Perron summation formula was introduced, and it was indicated how it works for such subblock counting problems. We will use this technique in the present paper. (A Delange type analysis would be feasible but very messy.)

Recently, Heuberger and Prodinger [10] have considered a symmetric system

$$n = \sum_{j=0}^{\infty} \varepsilon_j q^j$$

Date: October 28, 2002.
1991 *Mathematics Subject Classification.* Primary:11K16 ; Secondary:11A63, 11M41.
Key words and phrases. Digit representation, subblock occurrence, Mellin-Perron formula.
† This author is supported by the START-project Y96-MAT of the Austrian Science Fund.
‡ This author is supported by grant S8307-MAT of the Austrian Science Fund.

with an even base q and digits $\varepsilon_j \in \{-\frac{q}{2}, \dots, \frac{q}{2}\}$. Such a system is a priori redundant because of the existence of both $\pm\frac{q}{2}$ but can be made unique by the condition that $|\varepsilon_j| = q/2$ implies $0 \le \operatorname{sign}(\varepsilon_j)\varepsilon_{j+1} < q/2$ (Equivalent conditions where discussed in [10]). We call this expansion the *symmetric signed digit expansion* of n and denote it by $(\dots\varepsilon_2(n)\varepsilon_1(n)\varepsilon_0(n))$. For $q = 2$, this system was already considered by Reitwiesner in a computer science context in [15].

In this paper we are addressing the subblock counting problem in such symmetric signed digit expansions.

Before we announce our principal findings, we need some notation.

If a block $\mathbf{b} = (b_s, \dots, b_0)$ is given, we denote its *value* by $\mathsf{value}(\mathbf{b}) = \sum_\ell b_\ell q^\ell$.

We also use Iverson's notation, popularized in [8]: $[P]$ is defined to be 1 if condition P is true, and 0 otherwise. With this notation we can count the number of subblock occurrences of $\mathbf{b}$ in (the symmetric signed digit expansions of) n via

$$s_{\mathbf{b}}(n) = \sum_{k=0}^{\infty} [(\varepsilon_{k+r-1}(n), \dots, \varepsilon_k(n)) = \mathbf{b}]\,. \tag{1.1}$$

We only consider *admissible* blocks $\mathbf{b}$: these blocks represent the number $\mathsf{value}(\mathbf{b})$ in the symmetric signed digit expansion. For interest we note that there are

$$\frac{2+q}{1+q}q^r - \frac{1}{1+q}(-1)^r$$

admissible blocks of length r; this was implicitly proved in [9].

We also use the decomposition of a real number x as $x = \lfloor x \rfloor + \{x\}$ with the *fractional part* $\{x\}$ satisfying $0 \le \{x\} < 1$.

As said before, we are going to study the quantity

$$S_{\mathbf{b}}(N) = \sum_{n<N} \sum_{k=0}^{\infty} [(\varepsilon_{k+r-1}(n), \dots, \varepsilon_k(n)) = \mathbf{b}]\,. \tag{1.2}$$

We will prove the following theorem.

Theorem 1. *Let $q \ge 2$ be an even integer and $r \ge 1$. For an admissible block $\mathbf{b} = (b_{r-1}, \dots, b_0)$ with $|b_{r-1}| < \frac{q}{2}$ and $\mathbf{b} \ne 0^r$ the number of occurrences of the block $\mathbf{b}$ in the symmetric signed digit expansions of the positive integers less than N satisfies*

$$S_{\mathbf{b}}(N) = \frac{Q(b_0)}{q^r(q+1)} N \log_q N + h_0(\mathbf{b})N + N H_{\mathbf{b}}(\log_q N) + o(N), \tag{1.3}$$

where

$$Q(\eta) = q + \begin{cases} 2 & \text{for } \eta = 0, \\ 0 & \text{for } \eta = \pm\frac{q}{2}, \\ 1 & \text{else,} \end{cases} \tag{1.4}$$

$$H_{\mathbf{b}}(x) = \sum_{k \in \mathbb{Z}\setminus\{0\}} h_k(\mathbf{b}) e^{2k\pi i x},$$

$$\begin{aligned}
h_k(\mathbf{b}) =&\frac{\log q}{2k\pi i(\log q+2k\pi i)}\left(\zeta\left(\frac{2k\pi i}{\log q},[\mathsf{value}(\mathbf{b})<0]+q^{-r}\,\mathsf{value}(\mathbf{b})+\frac{R_{\min}(b_0)}{q^r(q+1)}\right)\right.\\
&\left.-\zeta\left(\frac{2k\pi i}{\log q},[\mathsf{value}(\mathbf{b})<0]+q^{-r}\,\mathsf{value}(\mathbf{b})+\frac{R_{\max}(b_0)}{q^r(q+1)}\right)\right) \quad \textit{for } k\neq 0,\\
h_0(\mathbf{b}) =&\log_q\Gamma\left([\mathsf{value}(\mathbf{b})<0]+q^{-r}\,\mathsf{value}(\mathbf{b})+\frac{R_{\min}(b_0)}{q^r(q+1)}\right)\\
&-\log_q\Gamma\left([\mathsf{value}(\mathbf{b})<0]+q^{-r}\,\mathsf{value}(\mathbf{b})+\frac{R_{\max}(b_0)}{q^r(q+1)}\right)\\
&-\frac{Q(b_0)}{q^r(q+1)}\left(r+\frac{1}{2}+\frac{1}{\log q}-\frac{1}{q+1}\right)+\frac{1}{q^{r-1}(q+1)},
\end{aligned}$$

$$R_{\min}(\eta) = -\frac{q}{2}-\left[(\eta-1)\bmod q\geq\frac{q}{2}\right], \tag{1.5}$$

$$R_{\max}(\eta) = \frac{q}{2}+\left[\eta\bmod q<\frac{q}{2}\right]. \tag{1.6}$$

The function $H_{\mathbf{b}}(x)$ is a periodic continuous function of period 1 and mean 0. As usual $\zeta(s,x)$ denotes the Hurwitz ζ-function.

Remark. *The case of blocks $\mathbf{b}$ with most significant digit $b_{r-1}=\pm\frac{q}{2}$ can be reduced to Theorem 1 by using the following simple observation*

$$S_{\mathbf{b}}(N)=\begin{cases}\displaystyle\sum_{\eta=-\frac{q}{2}+1}^{0}S_{\eta\mathbf{b}}(N) & \textit{for } b_{r-1}=-\frac{q}{2},\\ \displaystyle\sum_{\eta=0}^{\frac{q}{2}-1}S_{\eta\mathbf{b}}(N) & \textit{for } b_{r-1}=\frac{q}{2}.\end{cases}$$

The main term in this case is $\frac{1}{2}\frac{Q(b_0)}{q^r(q+1)}N\log_q N$.

The instance $r=1$ (counting digits) was discussed in [10], although without mentioning the periodic fluctuations in explicit form. Thuswaldner [17] has used Dirichlet series and the Mellin-Perron summation formula to exhibit this fluctuating behaviour in the case $q=2$ and $r=1$.

The limit distribution of digital functions of various kinds has been investigated by several authors. Especially, we mention the work of M. Drmota and J. Gajdosik [4] for local and central limit theorems for the sum-of-digits function with respect to recurrence based numeration systems. Furthermore, central limit theorems for digital functions in polynomial subsequences of the integers have been studied by N. L. Bassily and I. Katai [1, 2].

We will prove the following central limit theorem for $s_{\mathbf{b}}(n)$.

Theorem 2. *Let $s_{\mathbf{b}}(n)$ denote the number of occurrences of the block* $\mathbf{b}$ *in the symmetric signed digit expansion of n defined in* (1.1). *Then $s_{\mathbf{b}}$ satisfies*

$$\lim_{N\to\infty}\frac{1}{N}\#\left\{n<N \mid s_{\mathbf{b}}(n)<\frac{Q(b_0)}{q^r(q+1)}\log_q N + x\sqrt{V_{\mathbf{b}}\log_q N}\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt, \tag{1.7}$$

where

$$\begin{aligned} V_{\mathbf{b}} = \frac{Q(b_0)}{q^r(q+1)}\Bigg(1+2\sum_{t=1}^{r-1}[b_t=b_0,\dots,b_{r-1}=b_{r-1-t}]\frac{1}{q^t} \\ +2\frac{q}{q^r(q+1)}\left(1-\frac{Q(b_0)}{q+1}\right)-(2r-1)\frac{Q(b_0)}{q^r(q+1)}\Bigg). \end{aligned} \tag{1.8}$$

2. Explicit formulæ

In [10] we were able to give an explicit formula for the k-th digit of the symmetric signed digit expansion of n. The problem in this section is to combine this information for each individual digit of a given admissible block $\mathbf{b}$ in a manageable way.

Theorem 5 of [10] asserts that the digits of the symmetric signed digit expansion can be computed by

$$\varepsilon_k(n) = \sum_{s=0}^{q^2-1} c_s\left\lfloor\frac{n}{q^{k+2}}+\xi_s\right\rfloor, \tag{2.1}$$

where $k\geq 0$ and

$$\begin{aligned} \xi_s &:= \frac{s(q+1)+q/2+[s \bmod q<q/2]}{q^2(q+1)},\\ c_s &:= \begin{cases} -(q-1) & \text{if } s \bmod q = q/2-[\lfloor s/q\rfloor\geq q/2],\\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

for $0\leq s<q^2$.

It will be convenient to extend (2.1) to arbitrary reals x (instead of the integer n) and arbitrary integers k. We now show that this indeed defines a digit expansion for every real x.

In the proof of Theorem 5 in [10], (2.1) has been rewritten to

$$\varepsilon_k(x)=\sum_{j=0}^{q-1}\left(\left\lfloor\frac{x}{q^{k+1}}+\frac{j}{q}+\frac{q/2+[j<q/2]}{q(q+1)}\right\rfloor - q\left\lfloor\frac{x}{q^{k+2}}+\frac{j}{q}+\frac{q/2+[j<q/2]}{q(q+1)}\right\rfloor\right).$$

It is clear that $\varepsilon_k(x)=0$ for sufficiently large k. Therefore, we obtain

$$\sum_{k=-L}^{\infty}\varepsilon_k(x)q^k=\sum_{j=0}^{q-1}q^{-L}\left\lfloor xq^{L-1}+\frac{j}{q}+\frac{q/2+[j<q/2]}{q(q+1)}\right\rfloor = x+\mathcal{O}(q^{-L}).$$

This implies $\sum_{k\in\mathbb{Z}} \varepsilon_k(x)q^k = x$ for all $x \in \mathbb{R}$.

Furthermore, the proof of Theorem 5 in [10] shows that $|\varepsilon_k(x)| \leq \frac{q}{2}$ and that $|\varepsilon_k(x)| = \frac{q}{2}$ implies $0 \leq \operatorname{sign}(\varepsilon_k(x))\varepsilon_{k+1}(x) \leq \frac{q}{2} - 1$.

We also state (2.1) in an alternative way:

Lemma 1. *Let $k \in \mathbb{Z}$, $x \in \mathbb{R}$ and let $\xi_{-1} := 0$ and $\xi_{q^2} := 1$. Then the ξ_ℓ are increasing. Choose $0 \leq s \leq q^2$ such that*

$$\xi_{s-1} \leq \left\{\frac{x}{q^{k+2}}\right\} < \xi_s. \tag{2.2}$$

Write $s = mq + j$ where $0 \leq m$ and $0 \leq j < q$.

Then the kth digit $\varepsilon_k(n)$ of n can be expressed as

$$\varepsilon_k(n) = \varepsilon_0(s) = \begin{cases} j & \text{if } j < q/2, \\ j - q & \text{if } j > q/2, \\ q/2 & \text{if } j = q/2 \text{ and } m < q/2, \\ -q/2 & \text{if } j = q/2 \text{ and } m \geq q/2. \end{cases} \tag{2.3}$$

Proof. It can easily be checked that the ξ_ℓ are increasing. Since $\sum_{\ell=0}^{q^2-1} c_\ell = 0$ and since $\xi_\ell + \xi_{q^2-1-\ell} = 1$ for $-1 \leq \ell \leq q^2$, we can rewrite (2.1) as

$$\varepsilon_k(x) = \sum_{\ell=0}^{q^2-1} c_\ell \left\lfloor y + 1 - \xi_{q^2-1-\ell} \right\rfloor,$$

where $y = \left\{xq^{-(k+2)}\right\}$. By the monotonicity of the ξ_ℓ, relation (2.2) implies $\lfloor y + 1 - \xi_\ell \rfloor = [\ell \leq s-1]$. Therefore, we have

$$\varepsilon_k(x) = \sum_{\ell=0}^{s-1} c_\ell = \sum_{t=0}^{m-1}\sum_{\ell=0}^{q} c_{tq+\ell} + \sum_{\ell=0}^{j-1} c_{mq+\ell}.$$

By definition, $\sum_{\ell=0}^{q} c_{tq+\ell} = 0$ for all $0 \leq t \leq q-1$. This implies that

$$\varepsilon_k(x) = \sum_{\ell=0}^{j-1} c_{mq+\ell} = j - q\left[j - 1 \geq q/2 - [m \geq q/2]\right].$$

This proves the first equation in (2.3). The second equation in (2.3) is Lemma 3 in [10]. $\square$

We will now find out how the block $(\varepsilon_{k+r-1}(n), \ldots, \varepsilon_k(n))$ can be calculated from the knowledge of $\left\{n/q^{k+r+1}\right\}$. To this aim, we fix some $r \geq 1$, $k \geq 0$ and some $1 \leq j \leq q^{r+1}(q+1)$. We consider an integer n such that

$$\left\{\frac{n}{q^{k+r+1}}\right\} \in \left[\frac{j-1}{q^{r+1}(q+1)}, \frac{j}{q^{r+1}(q+1)}\right). \tag{2.4}$$

This implies that

$$\frac{(j-1)/q^\ell}{q^2(q+1)} \leq \frac{n}{q^{k+\ell+2}} - uq^{r-\ell-1} < \frac{j/q^\ell}{q^2(q+1)}$$

for some integer u and all $0 \le \ell \le r-1$. By Lemma 1, the digit $\varepsilon_{k+\ell}(n)$ depends on j and ℓ only, and not on n: If $-1 \le s_\ell \le q^2$ is chosen such that

$$\xi_{s_\ell - 1} \le \left\{ \frac{j-1}{q^{\ell+2}(q+1)} \right\} < \xi_{s_\ell},$$

then

$$\varepsilon_{k+\ell}(n) = \varepsilon_0(s_\ell).$$

Using Lemma 1 once more, we get

$$\varepsilon_{k+\ell}(n) = \varepsilon_\ell\left(\frac{j-1}{q+1}\right). \tag{2.5}$$

Therefore, we study the digit expansion of real numbers in more detail.

Lemma 2. *For all $x \in \mathbb{R}$ there exist unique $u \in \mathbb{Z}$ and $v \in \mathbb{R}$ with $x = u + v$ such that*

$$-\frac{q/2 + [(\varepsilon_0(u) - 1) \bmod q \ge q/2]}{q+1} \le v < \frac{q/2 + [\varepsilon_0(u) \bmod q < q/2]}{q+1}. \tag{2.6}$$

Furthermore $\varepsilon_\ell(x) = \varepsilon_\ell(u)$ for $\ell \ge 0$.

Proof. Since

$$\frac{q/2 + [\varepsilon_0(u) \bmod q < q/2]}{q+1} = 1 - \frac{q/2 + [(\varepsilon_0(u+1) - 1) \bmod q \ge q/2]}{q+1}$$

existence and uniqueness of u and v follows.

We consider first the case $\ell = 0$. By Lemma 1, there is some s with $\varepsilon_0(u) = \varepsilon_0(s)$ and $\xi_{s-1} \le \{u/q^2\} < \xi_s$. We assume first $1 \le s < q^2$. By definition, we have

$$(s-1) + \frac{q/2 + [(s-1) \bmod q < q/2]}{q+1} \le u - mq^2 < s + \frac{q/2 + [s \bmod q < q/2]}{q+1}$$

for some integer m. Since u is an integer, we conclude that

$$(s-1) + \left\lceil \frac{q/2 + [(s-1) \bmod q < q/2]}{q+1} \right\rceil \le u - mq^2 \le s + \left\lceil \frac{q/2 + [s \bmod q < q/2]}{q+1} \right\rceil - 1.$$

Since the values of the ceiling functions equal 1, we can rewrite this as

$$\xi_{s-1} + \frac{q/2 + [(s-1) \bmod q \ge q/2]}{q^2(q+1)} \le \frac{u}{q^2} - m \le \xi_s - \frac{q/2 + [s \bmod q < q/2]}{q^2(q+1)}. \tag{2.7}$$

Combining this with (2.6) yields

$$\xi_{s-1} \le \frac{u+v}{q^2} - m < \xi_s,$$

which is equivalent to $\varepsilon_0(u+v) = \varepsilon_0(s) = \varepsilon_0(u)$.

For $s=0$ and $s=q^2$ we have to do some extra work. We only show what happens for $s=0$: we still have the upper bound as in (2.7). Thus the only problem occurs, if $\frac{u+v}{q^2}-m<0$. But then

$$\xi_{q^2}=1>\frac{u+v}{q^2}-m+1\geq 1-\frac{\frac{q}{2}+1}{q^2(q+1)}=\xi_{q^2-1}$$

by (2.6). This also implies $\varepsilon_0(x)=\varepsilon_0(u)$ as requested.

In order to deal with $\ell\geq 1$ we consider

$$\sum_{k\geq 0}(\varepsilon_k(x)-\varepsilon_k(u))q^k=v-\sum_{k<0}\varepsilon_k(x)q^k.$$

The left hand side is an integer which is divisible by q by the above discussion. The absolute value of the right hand side is at most

$$\frac{q/2+1}{q+1}+\frac{q}{2}\sum_{k\geq 1}q^{-k}<2\leq q,$$

which implies that both sides vanish. This means that $\sum_{k\geq 0}\varepsilon_k(x)q^k=\sum_{k\geq 0}\varepsilon_k(u)q^k$. Since both sides are symmetric signed digit expansions of the same integer, the digits have to be equal. □

3. Counting blocks

Let $\mathbf{b}=(b_{r-1},\dots,b_0)$ be an admissible block. We want to count the number of occurrences of the block $\mathbf{b}$ as a subblock of the digit expansions of the integers $0,\dots,N-1$. In order to avoid technical problems arising from leading zeros we exclude the block $\mathbf{b}=0^r$. Furthermore, we can exclude blocks $\mathbf{b}$ with most significant digit $b_{r-1}\in\{\pm\frac{q}{2}\}$. This ensures that $\varepsilon_\ell(wq^r+\mathsf{value}(\mathbf{b}))=b_\ell$ for $0\leq\ell\leq r-1$ and $w\in\mathbb{Z}$, because Algorithm 2 of [10] does not consider w when computing the first r digits in that case. For blocks starting with zeros, it makes sense to tacitly assume that any number has a sufficient number of leading zeros.

Let $\mathbf{b}$, $n\geq 1$ and $k\geq 0$ be fixed. By (2.4) and (2.5), we have $(\varepsilon_{k+r-1(n)},\dots,\varepsilon_k(n))=\mathbf{b}$ if and only if

$$\left\{\frac{n}{q^{k+r+1}}\right\}\in\bigcup_{j\in J(\mathbf{b})}\left[\frac{j-1}{q^{r+1}(q+1)},\frac{j}{q^{r+1}(q+1)}\right),$$

where

$$J(\mathbf{b})=\left\{1\leq j\leq q^{r+1}(q+1):\left(\varepsilon_{r-1}\left(\frac{j-1}{q+1}\right),\dots,\varepsilon_0\left(\frac{j-1}{q+1}\right)\right)=\mathbf{b}\right\}. \tag{3.1}$$

In order to describe the set $J(\mathbf{b})$, we write $j-1=(q+1)u+R$ with $R_{\min}(\varepsilon_0(u))\leq R<R_{\max}(\varepsilon_0(u))$ with $R_{\min}$ and $R_{\max}$ defined in (1.5). Then by Lemma 2 we have $\varepsilon_\ell(\frac{j-1}{q+1})=\varepsilon_\ell(u)$ for $\ell\geq 0$ and can rewrite (3.1) as

$$J(\mathbf{b})=\big\{j=(q+1)u+R+1:u\in\mathbb{Z},1\leq j\leq q^{r+1}(q+1),$$

$$R_{\min}(\varepsilon_0(u)) \le R < R_{\max}(\varepsilon_0(u)), \text{ and } (\varepsilon_{r-1}(u), \dots, \varepsilon_0(u)) = \mathbf{b}\bigr\}.$$

Under our assumptions on the block $\mathbf{b}$ we can write the numbers u satisfying the last condition as $u = wq^r + \mathsf{value}(\mathbf{b})$ with $w \in \mathbb{Z}$. Thus we arrive at

$$J(\mathbf{b}) = \bigl\{j = (q+1)(wq^r + \mathsf{value}(\mathbf{b})) + R + 1 : w \in \mathbb{Z}, 1 \le j \le q^{r+1}(q+1), \text{ and } R_{\min}(b_0) \le R < R_{\max}(b_0)\bigr\}.$$

Inserting the definition of j in the range given for j, we get the following condition on w:

$$-\left\lfloor q^{-r}\left(\mathsf{value}(\mathbf{b}) + \frac{R}{q+1}\right)\right\rfloor \le w < q - \left\lfloor q^{-r}\left(\mathsf{value}(\mathbf{b}) + \frac{R}{q+1}\right)\right\rfloor.$$

By (1.5) we have $\operatorname{sign}(\mathsf{value}(\mathbf{b}) + \frac{R}{q+1}) = \operatorname{sign} \mathsf{value}(\mathbf{b}) \neq 0$ and therefore

$$0 < (\operatorname{sign} \mathsf{value}(\mathbf{b}))\left(\mathsf{value}(\mathbf{b}) + \frac{R}{q+1}\right) < q^r.$$

Summing up this gives

$$\left\lfloor q^{-r}\left(\mathsf{value}(\mathbf{b}) + \frac{R}{q+1}\right)\right\rfloor = -\left[\mathsf{value}(\mathbf{b}) < 0\right].$$

Thus we reach the following explicit characterization:

Proposition 1. *Let $0^r \neq \mathbf{b}$ an admissible block with $|b_{r-1}| < q/2$, n be a positive integer and $k \ge 0$. Let*

$$I_{\mathbf{b}} := \bigcup_{[\mathsf{value}(\mathbf{b})<0] \le w < q + [\mathsf{value}(\mathbf{b})<0]} \ \bigcup_{R_{\min}(b_0) \le R < R_{\max}(b_0)} \left[\frac{wq^r(q+1) + (q+1)\,\mathsf{value}(\mathbf{b}) + R}{q^{r+1}(q+1)}, \frac{wq^r(q+1) + (q+1)\,\mathsf{value}(\mathbf{b}) + R + 1}{q^{r+1}(q+1)}\right).$$

Then $((\varepsilon_{k+r-1}(n), \dots, \varepsilon_k(n)) = \mathbf{b}$ if and only if

$$\left\{\frac{n}{q^{k+r+1}}\right\} \in I_{\mathbf{b}}.$$

We now study the sum (1.2). We denote the interval $[\frac{j-1}{q^{r+1}(q+1)}, \frac{j}{q^{r+1}(q+1)})$ by I_j and its characteristic function by 1_{I_j}. The above proposition shows that

$$S_{\mathbf{b}}(N) = \sum_{w=[\mathsf{value}(\mathbf{b})<0]}^{q+[\mathsf{value}(\mathbf{b})<0]-1} \ \sum_{R=R_{\min}(b_0)}^{R_{\max}(b_0)-1} S_{wq^r(q+1)+(q+1)\,\mathsf{value}(\mathbf{b})+R+1}(N), \tag{3.2}$$

where

$$S_j(N) = \sum_{n<N} \sum_{k=0}^{\infty} 1_{I_j}\left(\left\{\frac{n}{q^{k+r+1}}\right\}\right). \tag{3.3}$$

We note that $(q+1)(wq^r + \mathsf{value}(\mathbf{b})) + R + 1 \ge 2$ by (1.5).

4. Dirichlet series

In this section, we study the asymptotic behaviour of the sum $S_j(N)$ defined in (3.3), where $1 < j \leq q^{r+1}(q+1)$ using Dirichlet generating functions and the Mellin-Perron summation formula.

We rewrite $S_j(N)$ as

$$S_j(N) = \sum_{n=1}^{N}(N-n)\sum_{k=0}^{\infty}\left(1_{I_j}(\{nq^{-k-r-1}\}) - 1_{I_j}(\{(n-1)q^{-k-r-1}\})\right) \tag{4.1}$$

using Abel summation. It is clear that the difference in (4.1) only takes values in $\{0, \pm 1\}$. We now discuss for which n the non-zero values are taken.

The first term in the difference equals 1, if

$$n \equiv u \mod q^{k+r+1} \quad \text{for some } u \in \left\{\left\lceil\frac{(j-1)q^k}{q+1}\right\rceil, \dots, \left\lceil\frac{jq^k}{q+1}\right\rceil - 1\right\}.$$

Using a similar expression for the second term in the difference, we obtain

$$1_{I_j}(\{nq^{-k-r-1}\}) - 1_{I_j}(\{(n-1)q^{-k-r-1}\}) = \\ \left[n \equiv \left\lceil\frac{(j-1)q^k}{q+1}\right\rceil \mod q^{k+r+1}\right] - \left[n \equiv \left\lceil\frac{jq^k}{q+1}\right\rceil \mod q^{k+r+1}\right].$$

Now we can write the Dirichlet generating function of

$$\sum_{k=0}^{\infty}\left(1_{I_j}(\{nq^{-k-r-1}\}) - 1_{I_j}(\{(n-1)q^{-k-r-1}\})\right)$$

as

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{n^s}\sum_{k=0}^{\infty}&\left(1_{I_j}(\{nq^{-k-r-1}\}) - 1_{I_j}(\{(n-1)q^{-k-r-1}\})\right) \\ &= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\frac{1}{\left(q^{k+r+1}n + \left\lceil\frac{(j-1)q^k}{q+1}\right\rceil\right)^s} - \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\frac{1}{\left(q^{k+r+1}n + \left\lceil\frac{jq^k}{q+1}\right\rceil\right)^s}.\end{aligned} \tag{4.2}$$

It is clearly enough to study functions $\psi_j(s)$ defined by

$$\psi_j(s) = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\frac{1}{\left(q^{k+r+1}n + \left\lceil\frac{jq^k}{q+1}\right\rceil\right)^s} = \sum_{n=1}^{\infty}\frac{\lambda_j(n)}{n^s} = \sum_{n=1}^{\infty}\frac{\lambda_j^{(0)}(n)}{n^s} + \sum_{n=1}^{\infty}\frac{\lambda_j^{(1)}(n)}{n^s},$$

where $\lambda_j^{(\ell)}(n)$ denotes the contribution originating from the terms $k \equiv \ell \mod 2$ ($\ell = 0, 1$) in the first sum. The splitting of the sum is motivated by

$$\left\lceil\frac{jq^k}{q+1}\right\rceil = \frac{jq^k}{q+1} + \frac{j(-1)^{k+1} \bmod (q+1)}{q+1}.$$

This leads us to studying the functions

$$\sum_{n=1}^{\infty}\frac{\lambda_j^{(\ell)}(n)}{(n-\alpha_j^{(\ell)})^s}=\sum_{\substack{k\geq 0\\ k\equiv \ell \bmod 2}}\sum_{n=0}^{\infty}\frac{1}{\left(q^{k+r+1}n+\frac{jq^k}{q+1}\right)^s}=\frac{q^{-(r+\ell-1)s}}{q^{2s}-1}\zeta\left(s,\frac{j}{(q+1)q^{r+1}}\right), \tag{4.3}$$

where $\alpha_j^{(\ell)}=\frac{j(-1)^{\ell+1} \bmod (q+1)}{q+1}$.

We will use the Mellin-Perron summation formula (cf. [16]) in the form

$$\sum_{n<N}(N-n)a_n=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\sum_{n=1}^{\infty}\frac{a_n}{(n-\alpha)^s}(N-\alpha)^{s+1}\frac{ds}{s(s+1)}, \tag{4.4}$$

where $0\leq\alpha<1$ and c is in the half-plane of absolute convergence of the Dirichlet series. In the sequel we will write $\int_{(c)}$ for the contour integral over the vertical line $\Re s=c$. The usefulness of this in the context of digital counting was described (for $\alpha=0$) in the survey [6] and in the slightly more general situation $0\leq\alpha<1$ in [7]. Without this version with the parameter α, one could still proceed successfully, as in [17], but that would be considerably more cumbersome and less elegant.

Applying (4.4) to the two functions in (4.3) separately with the two values of $\alpha_j^{(\ell)}$ we obtain

$$\begin{aligned}\sum_{n<N}(N-n)\lambda_j(n)&=\sum_{n<N}(N-n)\lambda_j^{(0)}(n)+\sum_{n<N}(N-n)\lambda_j^{(1)}(n)\\&=\frac{1}{2\pi i}\int_{(2)}\frac{q^{-(r-1)s}}{q^{2s}-1}\zeta(s,\beta_j)(N-\alpha_j^{(0)})^{s+1}\frac{ds}{s(s+1)}\\&+\frac{1}{2\pi i}\int_{(2)}\frac{q^{-rs}}{q^{2s}-1}\zeta(s,\beta_j)(N-\alpha_j^{(1)})^{s+1}\frac{ds}{s(s+1)},\end{aligned} \tag{4.5}$$

where $\beta_j=\frac{j}{(q+1)q^{r+1}}$ We now notice that $\zeta(\sigma+it,\alpha)=\mathcal{O}(|t|^{\frac12-\sigma})$ for $\sigma\leq 0$ (cf. [16]). Thus we can shift the line of integration to $\Re s=-\frac14$ by taking residues at the poles in $s=1$

and $s = \frac{k\pi i}{\log q} =: \chi_k$ into account to obtain
(4.6)
$$\begin{aligned}\sum_{n<N}(N-n)\lambda_j(n) &= \frac{q^{-(r-1)}}{2(q^2-1)}(N-\alpha_j^{(0)})^2 + \frac{q^{-r}}{2(q^2-1)}(N-\alpha_j^{(1)})^2\\
&+ \frac{1-2\beta_j}{4}(N-\alpha_j^{(0)})\log_q(N-\alpha_j^{(0)}) + \frac{1-2\beta_j}{4}(N-\alpha_j^{(1)})\log_q(N-\alpha_j^{(1)})\\
&- \frac{1-2\beta_j}{4}\left(r+\frac{1}{\log q}\right)(N-\alpha_j^{(0)}) - \frac{1-2\beta_j}{4}\left(r+1+\frac{1}{\log q}\right)(N-\alpha_j^{(1)})\\
&+ \frac{1}{2\log q}\zeta'(0,\beta_j)(2N-\alpha_j^{(0)}-\alpha_j^{(1)})\\
&+ (N-\alpha_j^{(0)})\sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{(-1)^{k(r-1)}}{2\log q}\frac{\zeta(\chi_k,\beta_j)}{\chi_k(\chi_k+1)}(N-\alpha_j^{(0)})^{\chi_k}\\
&+ (N-\alpha_j^{(1)})\sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{(-1)^{kr}}{2\log q}\frac{\zeta(\chi_k,\beta_j)}{\chi_k(\chi_k+1)}(N-\alpha_j^{(1)})^{\chi_k}\\
&+ \frac{1}{2\pi i}\int_{(-\frac{1}{4})}\cdots ds + \frac{1}{2\pi i}\int_{(-\frac{1}{4})}\cdots ds.\end{aligned}$$

It is clear from the absolute convergence of the two integrals that they are bounded by $\mathcal{O}(N^{\frac{3}{4}})$.

Similarly to the cases studied in [6] and [7] the above integrals could be computed by writing $\frac{1}{q^{2s}-1}$ as a geometric series and shifting the line of integration back to $\Re s = 2$. Since the resulting formulæ are highly unpleasant, we restrict ourselves to the two asymptotic main terms.

Thus we have

$$\begin{aligned}\sum_{n<N}(N-n)\lambda_j(n) &= \frac{1}{2}\frac{1}{q^r(q-1)}N^2 + \left(\frac{1}{2}-\beta_j\right)N\log_q N\\
&+ \left(\log_q\Gamma(\beta_j) - \frac{1}{2}\log_q 2\pi - \left(\frac{1}{2}-\beta_j\right)\left(\frac{1}{\log q}+\frac{1}{2}+r\right) - \frac{q\alpha_j^{(0)}+\alpha_j^{(1)}}{q^r(q^2-1)}\right)N\\
&+ N\left(F_j(\log_q N)+F_j(\log_q N+1)\right) + \mathcal{O}(\log N) + \mathcal{O}\Bigl(N\omega\Bigl(F_j,\frac{1}{N}\Bigr)\Bigr),\end{aligned} \tag{4.7}$$

where the periodic function F_j of period 2 is given by its absolutely convergent Fourier series

$$F_j(x) = \sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{(-1)^{kr}}{2\log q}\frac{\zeta(\chi_k,\beta_j)}{\chi_k(\chi_k+1)}e^{k\pi i x}$$

and $\omega(F_j,\delta)$ denotes the modulus of continuity of F_j. Clearly, $G_j(x)=F_j(x)+F_j(x+1)$ is a periodic function of period 1 with Fourier series

$$G_j(x)=\sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{1}{\log q}\frac{\zeta(\chi_{2k},\beta_j)}{\chi_{2k}(\chi_{2k}+1)}e^{2k\pi ix}.$$

Putting everything together we obtain

$$\begin{aligned}
S_j(N)=&\sum_{n<N}(N-n)(\lambda_{j-1}(n)-\lambda_j(n))\\
=&\frac{1}{(q+1)q^{r+1}}N\log_q N\\
&+\Bigg(\log_q\frac{\Gamma(\beta_{j-1})}{\Gamma(\beta_j)}-\frac{1}{(q+1)q^{r+1}}\left(\frac{1}{\log q}+\frac{1}{2}+r\right)-\frac{1}{q^r(q+1)^2}\\
&\qquad+\frac{q[j\equiv 1 \bmod (q+1)]-[j\equiv 0 \bmod (q+1)]}{q^r(q^2-1)}\Bigg)N\\
&+N\left(G_{j-1}(\log_q N)-G_j(\log_q N)\right)+o(N).
\end{aligned}\tag{4.8}$$

In order to obtain an asymptotic expression for $S_{\mathbf{b}}(N)$ stated in Theorem 1 we have to combine the equations (3.2) and (4.8). For this purpose we need two identities:

$$\sum_{w=0}^{q-1}\log_q\Gamma\left(x+\frac{w}{q}\right)=\frac{q-1}{2}\log_q 2\pi+\frac{1}{2}-qx+\log_q\Gamma(qx)\quad\text{for } q\geq 2\quad\text{(cf. [13])},$$

$$\sum_{w=0}^{q-1}\zeta\left(s,\frac{w}{q}+x\right)=q^s\sum_{n=0}^{\infty}\sum_{w=0}^{q-1}\frac{1}{(nq+w+qx)^s}=q^s\zeta(s,qx)\quad\text{for } x>0.$$

Furthermore, we notice that $[\mathsf{value}(\mathbf{b})<0]+q^{-r}\,\mathsf{value}(\mathbf{b})+\frac{R_{\min}(b_0)}{q^r(q+1)}>0$. This yields (1.3).

5. A central limit theorem

In this section we will sketch a proof of Theorem 2.

We first construct a probability space (Kubilius model, cf. [5]), which will be used to approximate the values of $s_{\mathbf{b}}(n)$ by random variables. For this purpose we consider the infinite product space

$$\Omega=\left\{-\frac{q}{2},\dots,\frac{q}{2}\right\}^{\mathbb{N}_0}$$

equipped with a probability measure μ given on cylinder sets by

$$\mu\left(\{(\omega_j)_{j\geq 0}\mid \omega_j=a_j \text{ for } j\leq k-1\}\right)=\lim_{N\to\infty}\frac{1}{N}\#\left\{n<N\mid \varepsilon_{k-1}(n)=a_{k-1},\dots,\varepsilon_0(n)=a_0\right\}.$$

The limit exists, since the subsets of $\mathbb{N}$ given by fixing a finite number of digits correspond to residue classes modulo a power of q by Proposition 1. From this fact it also follows that (for $j_0<j_1<\dots<j_s$)

$$\frac{1}{N}\#\left\{n < N \mid \varepsilon_{j_s}(n) = a_{j_s}, \ldots, \varepsilon_{j_0}(n) = a_{j_0}\right\} = \mu\left(\{(\omega_j)_{j\geq 0} \mid \omega_{j_s} = a_{j_s}, \ldots, \omega_{j_0} = a_{j_0}\}\right) + \mathcal{O}\left(\frac{q^{j_s}}{N}\right). \tag{5.1}$$

We define random variables

$$X_k(\omega) = [(\omega_{k+r-1}, \ldots, \omega_k) = \mathbf{b}].$$

From the definition of μ and Proposition 1 it follows that

$$\mu\left(X_k = 1\right) = \frac{Q(b_0)}{q^r(q+1)} + \frac{(-1)^k}{q^{k+r}}\left(1 - \frac{Q(b_0)}{q+1}\right), \tag{5.2}$$

which shows that the random variables X_k are not quite identically distributed. Furthermore, we compute the joint distribution of X_k and X_{k+t} for $t > r$:

$$\mu\left(X_k = 1 \wedge X_{k+t} = 1\right) = \mu\left(X_k = 1\right)\left(\mu\left(X_{k+t} = 1\right) + \frac{(-1)^{t+r}}{q^t}\left(1 - \frac{(-1)^{k+r}}{q^{k+r}}\right)\left(1 - \frac{Q(b_0)}{q+1}\right)\right), \tag{5.3}$$

which shows that the random variables X_k and X_{k+t} are not independent. The probabilities in the case $t \leq r$ depend on the self-overlapping structure of $\mathbf{b}$ and can be computed using (5.2). The variance of $\sum_{k=0}^{K} X_k$ can be computed

$$\begin{aligned}\mathbb{V}\left(\sum_{k=0}^{K} X_k\right) = K\frac{Q(b_0)}{q^r(q+1)}&\left(1 + 2\sum_{t=1}^{r-1}[b_t = b_0, \ldots, b_{r-1} = b_{r-1-t}]\frac{1}{q^t}\right.\\ &\left.+ 2\frac{q}{q^r(q+1)}\left(1 - \frac{Q(b_0)}{q+1}\right) - (2r-1)\frac{Q(b_0)}{q^r(q+1)}\right) + \mathcal{O}(1) = V_{\mathbf{b}}K + \mathcal{O}(1).\end{aligned} \tag{5.4}$$

Equation (5.3) shows that the random variables X_k are φ-mixing in the sense of [14] with $\varphi(t) = \mathcal{O}(q^{-t})$. An application of [14, Theorem 1.2.3] yields the central limit theorem for X_k

$$\lim_{K\to\infty} \mu\left(\sum_{k=0}^{K} X_k < \frac{Q(b_0)}{q^r(q+1)}K + x\sqrt{KV_{\mathbf{b}}}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \tag{5.5}$$

For technical reasons we replace the function $s_{\mathbf{b}}(n)$ by

$$\tilde{s}_{\mathbf{b}}(n) = \sum_{k\leq \log_q N - \log^{\frac{1}{3}} N} [(\varepsilon_{k+r-1}(n), \ldots, \varepsilon_k(n)) = \mathbf{b}]$$

and notice that $s_{\mathbf{b}}(n) - \tilde{s}_{\mathbf{b}}(n) = \mathcal{O}(\log^{\frac{1}{3}} N)$. From (5.1) it follows that the normalized moments of $\tilde{s}_{\mathbf{b}}(n)$ converge to the same limit as the corresponding moments of

$\sum_{k\le\log_q N-\log^{\frac{1}{3}} N} X_k$. Since (5.5) holds the convergence of moments implies

$$\lim_{N\to\infty}\frac{1}{N}\#\left\{n<N \mid \tilde{s}_{\mathbf{b}}(n)<\frac{Q(b_0)}{q^r(q+1)}\log_q N+x\sqrt{V_{\mathbf{b}}\log_q N}\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\,dt$$

by the Fréchet-Shohat theorem (cf. [5, Lemma 1.43]).

Acknowledgement. The authors thank an anonymous referee for pointing out the possiblity to prove a central limit theorem.

References

[1] N. L. Bassily. Distribution of q-ary digits in some sequences of integers. *Ann. Univ. Sci. Budapest. Sect. Comput.*, 14:13–22, 1994. Festschrift for the 50th birthday of Karl-Heinz Indlekofer.

[2] N. L. Bassily and I. Kátai. Distribution of the values of q-additive functions on polynomial sequences. *Acta Math. Hungar.*, 68:353–361, 1995.

[3] H. Delange. Sur la fonction sommatoire de la fonction"somme des chiffres". *Enseignement Math. (2)*, 21(1):31–47, 1975.

[4] M. Drmota and J. Gajdosik. The distribution of the sum-of-digits function. *J. Théor. Nombres Bordeaux*, 10:17–32, 1998.

[5] P. D. T. A. Elliott. *Probabilistic number theory. I, Mean-value theorems*, volume 239 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, New York, 1979.

[6] Ph. Flajolet, P. J. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy. Mellin transforms and asymptotics: digital sums. *Theoret. Comput. Sci.*, 123:291–314, 1994.

[7] P. J. Grabner, P. Kirschenhofer, and H. Prodinger. The sum-of-digits function for complex bases. *J. London Math. Soc. (2)*, 57(1):20–40, 1998.

[8] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics. A foundation for computer science.* Addison-Wesley, second edition, 1994.

[9] C. Heuberger and H. Prodinger. Carry propagation in signed digit representations. Preprint. Available at `http://www.opt.math.tu-graz.ac.at/~cheub/publications/carry.ps`.

[10] C. Heuberger and H. Prodinger. On minimal expansions in redundant number systems: Algorithms and quantitative analysis. *Computing*, 66:377–393, 2001.

[11] P. Kirschenhofer. Subblock occurrences in the q-ary representation of n. *SIAM J. Algebraic Discrete Methods*, 4(2):231–236, 1983.

[12] P. Kirschenhofer and H. Prodinger. Subblock occurrences in positional number systems and Gray code representation. *J. Inform. Optim. Sci.*, 5(1):29–42, 1984.

[13] W. Magnus, F. Oberhettinger, and R. P. Soni. *Formulas and Theorems for the Special Functions of Mathematical Physics.* Springer, Berlin, New York, 1966.

[14] W. Philipp. *Mixing sequences of random variables and probabilistic number theory.* Number 114 in Memoirs of the American Mathematical Society. American Mathematical Society, Providence, R. I., 1971.

[15] G. W. Reitwiesner. Binary arithmetic. In *Advances in computers*, volume 1, pages 231–308. Academic Press, New York, 1960.

[16] G. Tenenbaum. *Introduction to analytic and probabilistic number theory.* Cambridge University Press, Cambridge, 1995.

[17] J. M. Thuswaldner. Summatory functions of digital sums occurring in cryptography. *Period. Math. Hungar.*, 38:111–130, 1999.

(P. Grabner) Institut für Mathematik A, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria
E-mail address: `peter.grabner@tugraz.at`

(C. Heuberger) Institut für Mathematik B, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria
E-mail address: `clemens.heuberger@tugraz.at`

(H. Prodinger) The John Knopfmacher Centre for Applicable Analysis and Number Theory, School of Mathematics, University of the Witwatersrand, P. O. Wits, 2050 Johannesburg, South Africa
E-mail address: `helmut@staff.ms.wits.ac.za`