ON ZECKENDORF AND BASE b DIGIT SUMS

CHRISTIAN BALLOT

ABSTRACT. J. Pihko presented an elementary proof of the fact that the average number of summands in the Zeckendorf representation of an integer n is asymptotically equal to $C \log n$ for some explicit constant C. We retain the central idea of that proof, but provide a new elementary method that has the advantage of being more concise, and to also explain the asymptotics of the average sum of digits of integers in base b.

1. Introduction

The purpose of this note is to present a self-contained, elementary and nearly common proof of the next two theorems.

Theorem 1.1. *As n tends to infinity, we have*

$$S_F(n) \sim c_F\, n \log n,$$

where $c_F = (\alpha - 1)/(\sqrt{5} \log \alpha)$.

Theorem 1.2. *As n tends to infinity, we have*

$$S_b(n) \sim c_b\, n \log n,$$

where $c_b = (b - 1)/(2 \log b)$.

We first explain the notation.

The Fibonacci sequence $(F_k)_{k \geq 0}$ is defined by $F_0 = 0$, $F_1 = 1$, and $F_{k+2} = F_{k+1} + F_k$ for all $k \geq 0$. The Zeckendorf representation [10] of a non-negative integer n is the only representation of n as a sum of distinct, non-consecutive Fibonacci numbers all of indices at least 2[1]. For example, $17 = 13 + 3 + 1 = F_7 + F_4 + F_2$. If $F_k \leq n < F_{k+1}$, then the largest summand in the Zeckendorf representation of n is necessarily F_k. We define $s_F(n)$ as the number of summands of the Zeckendorf representation of n. Thus, $s_F(17) = 3$. The function $S_F(n)$ is the sum of all $s_F(k)$ as k varies from 0 to n. Since $1 = F_2$, $2 = F_3$, and $3 = F_4$, we have, for instance, $S_F(3) = 3$. Here α is $(1 + \sqrt{5})/2$.

Similarly, given a base b, i.e., an integer $b \geq 2$, every non-negative integer n is well-known to be representable in unique fashion as a sum $\sum_{i=0}^{k} d_i b^i$, where the digits d_i satisfy $0 \leq d_i < b$ and $d_k \neq 0$ if $b^k \leq n < b^{k+1}$. Then $s_b(n) = \sum_{i=0}^{k} d_i$ and $S_b(n) = \sum_{j=1}^{n} s_b(j)$.

We briefly remark that $c_F \simeq 0.574$, $c_2 \simeq 0.721$, $c_3 \simeq 0.910$ and $c_4 \simeq 1.082$, which raises the question of the existence of comparable number systems that would be more economical than the Zeckendorf representation in terms of digit sum averages.

Theorem 1.2 was first stated in [2], where a concise elementary proof is presented. The case $b = 2$ of Theorem 1.2 appears in [1]. The referee points out that another elementary proof of Theorem 1.2, essentially due to G. Grekos, was given in [8]. The asymptotics of the functions S_F and S_b are known with greater precision than Theorems 1.1 and 1.2. In particular, the

[1]By convention the empty sum, i.e., a sum with no summands, is equal to 0.

differences $D_b(n) := S_b(n) - c_b n \log n$ and $D_F(n) := S_F(n) - c_F n \log n$ are known to be $O(n)$. For $D_b(n)$, this was shown in [6] and also in [3], although this result can be seen directly from the proof given in [2]. Section 3 of this note shows that our method also yields the result for both D_b and D_F. However, more is known since $D_b(n)$ and $D_F(n)$ are expressible explicitly in terms of continuous, nowhere differentiable functions, or nearly so in the case of D_F. This was carried out first for $b = 2$ in [9], for a general base in [5], and for the Fibonacci case in [4].

In the paper [7], Pihko reports an elementary proof of Theorem 1.1. The main idea is the introduction of the function R_F which we recall here.

Definition 1.3. *Suppose $F_k \le n < F_{k+1}$ for some $k \ge 0$. Then*

$$R_F(n) := \begin{cases} S_F(n), & \text{if } n = F_k; \\ R_F(F_k) + R_F(n - F_k), & \text{otherwise.} \end{cases}$$

If $n = \sum_{j=1}^{s} F_{i_j}$, where $i_{j+1} - i_j \ge 2$, $j = 1, \ldots, s-1$ and $i_1 \ge 2$, i.e., if $\sum_{j=1}^{s} F_{i_j}$ is the Zeckendorf representation of n, then, by iterating the definition of R_F, we find that $R_F(n) = \sum_{j=1}^{s} R_F(F_{i_j})$. Because the difference $S_F(n) - R_F(n)$ is $o(n \log n)$, showing that $R_F(n) \sim c_F n \log n$ entails Theorem 1.1. However, while reading [7], we found that some calculations and parts of the argument were long, sinuous and idiosyncratic. This brought about a more concise and direct proof, yet as elementary, which is reported herein. In our view, this note is a completion of the idea of [7]. Indeed, we use the R_F function introduced in [7], but use a Cesàro-like result, Theorem 1.4 below, that allows to move from the asymptotics of $R_F(F_k)/F_k$ to those of $R_F(n)/n$, for a general integer n, more directly.

Theorem 1.4. *Let $(u_n)_{n\ge0}$ and $(v_n)_{n\ge0}$ be sequences of non-negative real numbers with*

(i) *$v_n \sim a\rho^n$, for some $a > 0$ and some $\rho > 1$, and*
(ii) *$u_n \sim c\, n\, v_n$, for some $c > 0$, as n tends to infinity.*

Put $U_n = \sum_{k=0}^{n} \epsilon_k u_k$ and $V_n = \sum_{k=0}^{n} \epsilon_k v_k$, where $\epsilon_k \in \{0, 1, \ldots, B\}$, $k = 0, 1, \ldots, n$, $\epsilon_n \ne 0$ and B is an integer ≥ 2.

Then $U_n \sim c\, n\, V_n$ regardless of the choice of the ϵ_k's, $k = 0, 1, \ldots, n$.

Remark. It is a Cesàro-like theorem in the sense that $u_n/v_n \sim c\,n \implies U_n/V_n \sim c\,n$.

Then, in analogy with the R_F function and in order to tackle Theorem 1.2, we define the function R_b below.

Definition 1.5. *Let $R_b(0) = 0$. Suppose $b^k \le n < b^{k+1}$ for some $k \ge 0$. Then*

$$R_b(n) := \begin{cases} S_b(n), & \text{if } n = b^k; \\ R_b(b^k) + R_b(n - b^k), & \text{otherwise.} \end{cases}$$

By iterating the above definition, we find that $R_b(n) = \sum_{i=0}^{k} d_i R_b(b^i)$, if $n = \sum_{i=0}^{k} d_i b^i$, $0 \le d_i < b$.

For example, suppose $b = 3$ and $n = 47$. Then $n = 3^3 + 2 \cdot 3^2 + 2 \cdot 3^0$ and $R_3(n) = R_3(3^3) + R_3(20) = R_3(3^3) + R_3(3^2) + R_3(11) = R_3(3^3) + R_3(3^2) + R_3(3^2) + R_3(2) = R_3(3^3) + 2 \cdot R_3(3^2) + 2 \cdot R_3(3^0)$.

Again, the asymptotic behavior of $S_b(n)/n$ for general n is derived from that of $S_b(b^k)/b^k$ using Theorem 1.4.

2. Proof of the Theorems

Proof of Theorem 1.4. Note that $\rho^n \ll V_n$ and $V_n \ll \rho^n$. Indeed, $V_n \geq v_n = a\rho^n\big(1+o(1)\big)$ and

$$V_n \leq B\sum_{k=0}^{n} v_k \sim aB\sum_{k=0}^{n}\rho^k \sim \frac{aB}{\rho-1}\rho\cdot\rho^n.$$

Let us fix a positive ε. By hypothesis, there exists a k_0 such that $k > k_0$ implies that $|u_k - ckv_k| \leq \frac{1}{3}\varepsilon k v_k$. Hence,

$$\begin{aligned}\left|\frac{U_n}{nV_n} - c\right| &= (nV_n)^{-1}\left|\sum_{k=0}^{n}\epsilon_k u_k - cn\sum_{k=0}^{n}\epsilon_k v_k\right| \\ &\leq (nV_n)^{-1}\sum_{k=0}^{n}\epsilon_k|u_k - cnv_k| \\ &\leq (nV_n)^{-1}\left[\sum_{k=0}^{n}\epsilon_k|u_k - ckv_k| + c\sum_{k=0}^{n}\epsilon_k v_k(n-k)\right].\end{aligned}$$

We break the sum $\sum_{k=0}^{n}\epsilon_k|u_k - ckv_k|$ into two subsums S_1 and S_2, where $S_1 := \sum_{k=0}^{k_0}\epsilon_k|u_k - ckv_k|$ is bounded above by $B\sum_{k=0}^{k_0}|u_k - ckv_k|$, a constant independent of n, and

$$\begin{aligned}S_2 := \sum_{k=k_0+1}^{n}\epsilon_k|u_k - ckv_k| &\leq \frac{\varepsilon}{3}\sum_{k_0+1}^{n}\epsilon_k k v_k \\ \leq \frac{\varepsilon}{3}\sum_{0}^{n}\epsilon_k k v_k \leq \frac{\varepsilon}{3}n\sum_{0}^{n}\epsilon_k v_k &= \frac{\epsilon}{3}nV_n.\end{aligned}$$

Thus, for n large enough, $S_1/(nV_n) \leq \varepsilon/3$ and $S_2/(nV_n) \leq \varepsilon/3$.

It remains to show that the sum $S_3 := \sum_{k=0}^{n}\epsilon_k v_k(n-k)$ is $o(nV_n)$.

Note that S_3 will be $o(nV_n)$ if the sum $\sum_{k=0}^{n-1}v_k(n-k)$ is $o(nV_n)$, or equivalently, if $\sum_{k=0}^{n-1}\rho^k(n-k)$ is $o(n\rho^n)$, i.e., by a change of variable, if $\sum_{k=1}^{n}k\rho^{n-k}$ is $o(n\rho^n)$. But clearly that last statement is true for

$$\sum_{k=1}^{n}k\rho^{n-k} < \rho^n\sum_{k=1}^{\infty}k\rho^{-k} \ll \rho^n.$$

Therefore, for n large enough, we also have $S_3/(nV_n) \leq \varepsilon/3$, and we conclude that $\left|\frac{U_n}{nV_n} - c\right| \leq \varepsilon$ for such n's. □

2.1. **Proof of Theorem 1.1.**

We recall that, by running the recursion $F_{k+2} = F_{k+1} + F_k$ backward, Fibonacci numbers may be defined for negative integers k. We then get $F_{-k} = (-1)^{k+1}F_k$ for all k. Also $F_k = (\alpha^k - \beta^k)/\sqrt{5}$, where $\beta = (1-\sqrt{5})/2$. The kth Lucas number is $L_k = \alpha^k + \beta^k$. We lay stress on the facts that $1 < \alpha < 2$ and that $-1 < \beta < 0$.

Lemma 2.1. *We have, for $k \geq 0$,*

$$S_F(F_{k+1}) = 1 + \frac{1}{5}\big[(k-1)L_k + 2F_{k-1}\big]. \tag{2.1}$$

Thus, for integers n of the form F_k, we have $S_F(n) \sim c_F n\log n$ with $c_F = \frac{\alpha-1}{\sqrt{5}\log\alpha}$.

Proof. Note that the sequence $(1/5)\big((k-1)L_k + 2F_{k-1}\big)$ has initial values $0, 0, 1, 2$ for $k = 0, 1, 2$ and 3, and is linear recurring of order four since it is annihilated by the polynomial $(x-\alpha)^2(x-\beta)^2 = (x^2-x-1)^2$. (Here, we use the well-known facts, which are easy to check, that the sequences $(r^k)_k$ and $(kr^k)_k$ are respectively annihilated by $x-r$ and by $(x-r)^2$, and, thus, that both are annihilated by $(x-r)^2$. The polynomial action we refer to is defined via the right shift. That is, $x \cdot (u_k) = (u_{k+1})$. Thus, for instance, $x^2 - x - 1 \cdot (F_k) = 0$.) But the sequence (a_k), where $a_k = S_F(F_{k+1}-1)$, shares the same four initial values and follows the same order four recursion. Indeed, if $F_k \le \ell < F_{k+1}$, then $s_F(\ell) = 1 + s_F(\ell - F_k)$ and $0 \le \ell - F_k < F_{k-1}$. Hence, for $k \ge 2$,

$$S_F(F_{k+1}-1) - S_F(F_k - 1) = F_{k-1} + S_F(F_{k-1}-1).$$

That is, $a_k - a_{k-1} - a_{k-2} = F_{k-1}$, which implies that $(x^2-x-1)^2$ annihilates the sequence (a_k).

Now, since $F_k \sim \alpha^k/\sqrt{5}$, we find using (2.1) that

$$S_F(F_k) \sim (1/5)kL_{k-1} \sim \frac{\log F_k}{5\log\alpha}\alpha^{k-1} \sim \frac{1}{\alpha}\frac{\log F_k}{\sqrt{5}\log\alpha}F_k = c_F n \log n.$$

□

Remark. The pretty formula (2.1) appeared on page 44 of [8] and was given another proof. In [8, p. 43], we also find that $S(F_{k+1}-1) = \sum_{j=0}^{k} F_j F_{k-j}$. Note that the sequence $b_k := (1/5)\big((k-1)L_k + 2F_{k-1}\big)$ is also equal to $(1/5)\left((k-1)F_k + 2kF_{k-1}\right)$, and that $(b_k)_{k\ge -1}$ is the fundamental sequence with initial values $0, 0, 0, 1$ associated with $(x^2-x-1)^2$.

The next easy lemma was stated in [7] without proof. We provide one for the sake of comparison with the base b case.

Lemma 2.2. *If $F_k \le n < F_{k+1}$ for some $k \ge 2$, then*

$$0 \le S_F(n) - R_F(n) < F_k.$$

Proof. We proceed by strong induction on k. Note that the lemma holds for $n < 3 = F_4$. Thus, assume $k \ge 4$ and the lemma holds for all $n < F_k$.

If $F_k \le n < F_{k+1}$, then

$$\begin{aligned} S_F(n) - R_F(n) &= \big(S_F(F_k) + n - F_k + S_F(n-F_k)\big) - \big(R_F(F_k) + R_F(n-F_k)\big) \\ &= (n - F_k) + \big(S_F(n-F_k) - R_F(n-F_k)\big). \end{aligned}$$

By the inductive hypothesis, we get

$$0 \le S_F(n) - R_F(n) \le n - F_k + F_{k-2} < F_{k+1} - F_k + F_{k-2} = F_k.$$

□

Proof of Theorem 1.1. By Lemma 2.2, we have, for all $n \ge 1$, $0 \le S_F(n) - R_F(n) < n$. Hence, $S_F(n) \sim c_F n \log n$ if, and only if $R_F(n) \sim c_F n \log n$.

Let $v_k = F_k$ and $u_k = R_F(F_k)$. By Lemma 2.1, $u_k \sim c_F F_k \log(F_k)$. But $c_F F_k \log(F_k) \sim ckv_k$, with $c = c_F \log\alpha$. Therefore, by and with the notation of Theorem 1.4, we have that $U_k \sim ckV_k$ as $k \to \infty$, where, if $F_k \le n < F_{k+1}$, we have chosen the ϵ_i's such that $n = V_k = \sum_{i=0}^{k} \epsilon_i F_i$ is the Zeckendorf representation of n. In particular, $\epsilon_i \in \{0, 1\}$, $\epsilon_0 = \epsilon_1 = 0$ and $\epsilon_k = 1$. Then $U_k = \sum_{i=0}^{k} \epsilon_i R_F(F_i) = R_F(n)$. Therefore, $R_F(n) \sim (c_F \log\alpha)kn$. As $\alpha^k/\sqrt{5} \sim F_k \le n < F_{k+1}$, $\log n \sim k\log\alpha$. Hence, $R_F(n) \sim c_F n \log n$. □

2.2. **Proof of Theorem 1.2.** We imitate the approach of the previous subsection. However, to alleviate notation in subsequent proofs we write S instead of S_b and R in place of R_b.

Lemma 2.3.[2] *We have for all* $k \geq 0$

$$S_b(b^k) = 1 + \frac{(b-1)}{2} k b^k. \tag{2.2}$$

Thus, for integers n of the form b^k, we have $S_b(n) \sim c_b n \log n$ with $c_b = \frac{b-1}{2\log b}$.

Proof. If j and j' are integers in $[0, b^k)$ with $j + j' = b^k - 1$, then $s_b(j) + s_b(j') = k(b-1)$. If b is even, then there are precisely $b^k/2$ such pairs $\{j, j'\}$ with $j \neq j'$. Thus, (2.2) holds. If b is odd, there are $(b^k - 1)/2$ pairs $\{j, j'\}$, with $j \neq j'$ and $j + j' = b^k - 1$, and the singleton $(b^k - 1)/2$ whose sum of digits is $k(b-1)/2$. Thus, (2.2) holds again.

Now $S(b^k) = 1 + \frac{(b-1)}{2} k b^k \sim \frac{b-1}{2\log b} (\log b^k) b^k = c_b n \log n$. □

Lemma 2.4. *If $b^k \leq n < b^{k+1}$ for some $k \geq 0$, then*

$$0 \leq S_b(n) - R_b(n) < b^{k+3}.$$

Proof. We proceed by strong induction on k. If $k = 0$, then $1 \leq n < b$. So $R(n) = nR(1) = nS(1) = n = S(n)$ and $S(n) - R(n) = 0$. We now assume that $k \geq 1$ and that the lemma holds for all integers $< b^k$. Suppose $n = ib^k + m$ with $1 \leq i < b$ and $m < b^k$. Then

$$S(n) = S(ib^k - 1) + (m+1)i + S(m). \tag{2.3}$$

Indeed, if $(j-1)b^k \leq \ell < jb^k$, $1 \leq j < b$, then $s_b(\ell) = j - 1 + s_b(\ell - (j-1)b^k)$. Thus, for $1 \leq j \leq i$, $S(jb^k - 1) - S\big((j-1)b^k - 1\big) = b^k(j-1) + S(b^k - 1)$, where we have set $S(-1)$ equal to 0. Therefore,

$$S(ib^k - 1) = b^k \sum_{j=1}^{i} (j-1) + iS(b^k - 1). \tag{2.4}$$

Putting (2.3) and (2.4) together yields $S(n) = \frac{i(i-1)}{2} b^k + iS(b^k) + mi + S(m)$. Since $R(n) = iS(b^k) + R(m)$ and $m < b^k$, we get, using the inductive hypothesis, that

$$\begin{aligned} 0 \leq S(n) - R(n) &= \frac{i(i-1)}{2} b^k + mi + \big(S(m) - R(m)\big) \\ &< \frac{b^2}{2} b^k + b^{k+1} + b^{k+2} \leq \frac{b^{k+2}}{2} + \frac{b^{k+2}}{2} + b^{k+2} \leq b^{k+3}. \end{aligned}$$

□

Remark. By using sharper upper bounds in the last few steps of the proof, one can replace b^{k+3} by b^{k+2} in Lemma 2.4, and for, say $b = 2$, one can get down to b^{k+1}.

Proof of Theorem 1.2. By Lemma 2.4, we have $0 \leq S(n) - R(n) < b^3 n$. Thus, $R(n) - S(n) = O(n)$. Hence, $S(n) \sim c_b n \log n$ if, and only if $R(n) \sim c_b n \log n$.

Let $v_k = b^k$ and $u_k = R(b^k)$. By Lemma 2.3, $u_k \sim c_b b^k \log(b^k) = ckv_k$, with $c = (b-1)/2$. Thus, we may apply Theorem 1.4 to the sequences (u_k) and (v_k) and obtain, with notation from Theorem 1.4, that $U_k \sim ckV_k$. If $b^k \leq n < b^{k+1}$, then n is equal to some $V_k = \sum_{i=0}^{k} \epsilon_i b^i$, where $\epsilon_i \in \{0, 1, \ldots, b-1\}$ are the base b digits of n and $\epsilon_k \geq 1$. But then we have $R(n) = U_k$. Therefore, $R(n) \sim c_b n \log n$, as $k \sim \log n / \log b$. □

[2]This is Lemma 4.6 of [8], where it is proved by induction on k.

3. A Final Remark

If, in Theorem 1.4, we replace condition (ii) by the stronger condition that $u_n = cnv_n + O(v_n)$, say $|u_n - cnv_n| \le Kv_n$ for all n, then the conclusion of the theorem is also stronger, that is, $U_n = cnV_n + O(V_n)$.

Indeed, the proof of Theorem 1.4 now leads to

$$\left|\frac{U_n}{nV_n} - c\right| \ll \frac{1}{n},$$

because

$$S_1 + S_2 = \sum_{k=0}^{n} \epsilon_k |u_k - ckv_k| \le K \sum_{k=0}^{n} \epsilon_k v_k = KV_n, \text{ and}$$

$$S_3 = \sum_{k=0}^{n} v_k(n-k) \text{ was shown to be } \ll V_n.$$

This stronger condition is satisfied in both the Fibonacci and the base b cases. From equation (2.1), we have $R_F(F_k) - ckF_k = O(F_k)$, i.e, $u_k - ckv_k = O(v_k)$ in the notation of the proof of Theorem 1.1, while, in the base b case, we have $u_k - ckv_k = R_b(b^k) - ckb^k = O(1) = O(b^k)$ by (2.2). Therefore, $U_k - ckV_k = O(V_k)$ in both cases. Thus, we have $R_F(n) - c_F n \log n = O(n)$ and $R_b(n) - c_b n \log n = O(n)$. By Lemmas 2.2 and 2.4, $S_F(n) - R_F(n) = O(n)$ and $S_b(n) - R_b(n) = O(n)$. Hence, we get with little further expense stronger versions of Theorems 1.1 and 1.2, namely,

$$S_F(n) = c_F\, n \log n + O(n), \quad \text{and} \tag{3.1}$$

$$S_b(n) = c_b\, n \log n + O(n). \tag{3.2}$$

4. Acknowledgments.

We are thankful to the referee for a very careful reading of this note and to Jukka Pihko for scanning and sending pages of [8] from his home in Finland.

References

[1] R. Bellman and H. N. Shapiro, *On a problem in additive number theory*, Ann. of Math., **49.2** (1948), 333–340.
[2] L. E. Bush, *An asymptotic formula for the average sum of digits of integers*, American Math. Monthly, **47** (1940), 154–156.
[3] P. Cheo and S. Yien, *A problem on the k-adic representation of positive integers*, Acta Math. Sinica, **5** (1955), 433–438.
[4] J. Coquet and P. van den Bosch, *A summation formula involving Fibonacci digits*, J. Number Theory, **22** (1986), 139–146.
[5] H. Delange, *Sur la fonction sommatoire de la fonction 'Somme des Chiffres'*, Enseignement Math., **21** (1975), 31–47.
[6] L. Mirsky, *A theorem on representation of integers in the scale of r*, Scripta Math., **15** (1949), 11–12.
[7] J. Pihko, *On the average number of summands in the Zeckendorf representation*, Congr. Numer., **200** (2010), 317–324.
[8] J. Pihko, *An algorithm for the additive representation of positive integers*, Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes, No. 46 (1983), 54 pp.
[9] J. R. Trollope, *An explicit expression for binary digital sums*, Mathematics Magazine, **41** (1968), 21–25.
[10] E. Zeckendorf, *Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas*, Bull. Soc. Roy. Sci. Liège, **41** (1972), 179–182.

MSC2010: 11A63, 11B39

Département de Mathématiques et Mécanique, Université de Caen, F14032 Caen Cedex, France
E-mail address: christian.ballot@unicaen.fr