Journal of Integer Sequences, Vol. 12 (2009),
Article 09.5.6

Generating Functions for the Digital Sum and Other Digit Counting Sequences

Franklin T. Adams-Watters
FrankTAW@Netscape.net

Frank Ruskey
University of Victoria
Victoria, BC V8W 3P6
CANADA
ruskey@cs.uvic.ca

Abstract

A numeration system associates a unique string, $\Xi(n)$, with each positive integer n, where each string is over the same finite alphabet. Various digit counting statistics of $\Xi(n)$ are of interest with respect to a numeration system. An example is the digital sum, which is the sum of the digits in the number. We present a unifying framework for deriving identities for the generating functions of such statistics in many of the more popular numeration systems.

1 Introduction

Numeration systems provide a rich source of integer sequences. There are many interesting digit counting statistics that arise from the various numeration systems. A typical example, the digital sum, is explained below.

Given a number n represented in binary, $n = (b_d b_{d-1} \cdots b_1 b_0)_2$, the (binary) *digital sum* of n, denoted $s_2(n)$ is $b_0 + b_1 + \cdots + b_d$. The digital sum goes under several other names including sideways sum, sideways addition, population count, and Hamming weight. It is denoted νn in Knuth [5], but we use the notation used by Allouche and Shallit [1] (but due to earlier researchers; e.g., Bésineau [2] and Coquet and Toffin [3]). There are two natural ways in which we might extend the idea to k-ary numbers, either by summing digits, or by counting non-zeros. We use the notation $s_k(n)$ (again following [1]) for the sum-of-digits function and $c_k(n)$ for the counting non-zeros function.

According to OEIS A053735, the ordinary generating function of $s_k(n)$ has the beautiful expression given in Theorem 1 below. For $k = 2$ this generating function may be found in Knuth [5]; see exercise 7.1.3.41. The purpose of this short paper is to provide a proof of this and some other related generating functions that are in the OEIS — as part of a more generalized setting. Table 1 contains some of the sequences to which our results apply.

Theorem 1. *For all $k \geq 2$,*

$$\begin{aligned}\sum_{n\geq 0} s_k(n) z^n &= \frac{1}{1-z} \sum_{m\geq 0} \frac{z^{k^m} + 2z^{2k^m} + \cdots + (k-1)z^{(k-1)k^m}}{1 + z^{k^m} + z^{2k^m} + \cdots + z^{(k-1)k^m}} \\ &= \frac{1}{1-z} \sum_{m\geq 0} \frac{z^{k^m} - kz^{k^{m+1}} + (k-1)z^{(k+1)k^m}}{(1-z^{k^m})(1-z^{k^{m+1}})}.\end{aligned}$$

2 Numeration as a sequence of columns

Imagine a table comprised of infinite columns of numbers, $C_0, C_1, C_2, \ldots$. The numbers in each column are indexed starting at 0 and the numbers found in all of the columns all come from the same finite set. In a numeration system each row of the table is distinct. For example, in the binary number system, C_j consists of the periodic repetition of 2^j 0s followed by 2^j 1s.

What is the generating function function for the row sums of those columns? Suppose that the generating function for the m-th column is $C_m(z)$. Then

$$A(z) = \sum_{m\geq 0} C_m(z) \tag{1}$$

is the generating function for the row sums. That is, $\langle z^n \rangle A(z)$ is the sum of the numbers in the n-th row, where $\langle z^n \rangle$ means "coefficient of z^n". The generating function (1) will exist so long as there are constants c_n such that $\langle z^n \rangle C_m(z) = 0$ for all $m \geq c_n$. In many numeration systems the m-th column can be described as a infinite string of the form $\mathbf{s}_m \mathbf{t}_m^\infty$, where $\mathbf{s}_m$ and $\mathbf{t}_m$ are strings with $\mathbf{t}_m \neq \varepsilon$ and $\mathbf{t}_m^\infty$ denotes the infinite string $\mathbf{t}_m^\infty = \mathbf{t}_m \mathbf{t}_m \mathbf{t}_m \cdots$. Let $|\mathbf{s}|$ be the length of the string $\mathbf{s}$. If $S_m(z)$ and $T_m(z)$ are the generating functions (which are actually polynomials) of $\mathbf{s}$ and $\mathbf{t}$, respectively. Then

$$C_m(z) = S_m(z) + \frac{z^{|\mathbf{s}_m|} T_m(z)}{1 - z^{|\mathbf{t}_m|}} \tag{2}$$

Often $\mathbf{s}_m$ and $\mathbf{t}_m$ will have a special form that allows for further simplification of $C_m(z)$

In this paper the most general from that we use is shown below. Here the m-th column depends on integers b_m, a_m and u_m, and the sequence of numbers $\alpha_0, \alpha_1, \alpha_2, \ldots$. The following string is called the *column pattern*:

$$\mathbf{s}_m \mathbf{t}_m^\infty = 0^{b_m} (\alpha_0^{a_m} \alpha_1^{a_m} \cdots \alpha_{(u_m-1)}^{a_m})^\infty \tag{3}$$

For example, the column pattern for the binary digital sum is $(0^{2^m}1^{2^m})^\infty$; here $b_m = 0$, $a_m = 2^m$, $\alpha_0 = 1$, $\alpha_1 = 1$, and $u_m = 2$. The generating function of $\alpha_0^{a_m}\alpha_1^{a_m}\cdots\alpha_{(u_m-1)}^{a_m}$ is

$$A_m(z) = (\alpha_0 + \alpha_1 z^{a_m} + \alpha_1 z^{2a_m} + \cdots + \alpha_{u_m-1} z^{(u_m-1)a_m})\frac{1-z^{a_m}}{1-z}$$

Thus, by (2), the generating function for (3) is

$$\frac{z^{b_m}A_m(z)}{1-z^{u_m a_m}} = \frac{z^{b_m}}{1-z}\cdot\frac{\alpha_0 + \alpha_1 z^{a_m} + \alpha_2 z^{2a_m} + \cdots + \alpha_{u_m-1} z^{(u_m-1)a_m}}{1 + z^{a_m} + z^{2a_m} + \cdots + z^{(u_m-1)a_m}}.$$

Summing over $m \geq 0$ we obtain

$$A(z) = \frac{1}{1-z}\sum_{m\geq 0}\frac{z^{b_m}(1-z^{a_m})}{1-z^{u_m a_m}}(\alpha_0 + \alpha_1 z^{a_m} + \alpha_2 z^{2a_m} + \cdots + \alpha_{u_m-1} z^{(u_m-1)a_m}) \quad (4)$$

$$= \frac{1}{1-z}\sum_{m\geq 0} z^{b_m}\frac{\alpha_0 + \alpha_1 z^{a_m} + \alpha_2 z^{2a_m} + \cdots + \alpha_{u_m-1} z^{(u_m-1)a_m}}{1 + z^{a_m} + z^{2a_m} + \cdots + z^{(u_m-1)a_m}}. \quad (5)$$

In the sections to follow we apply this generating function to various numeration systems, starting with a new addition to the OEIS.

3 The Balanced Ternary System

In the *balanced ternary system* each natural number n is expressed as a sum of distinct signed powers of 3. For example $5 = 9 - 3 - 1 = 3^3 - 3^1 - 3^0$. The digital sum is OEIS A065363. Following Knuth [4] we use $\bar{1}$ to denote -1. It is outside the scope of this paper, but it is not difficult to show that the pattern of the m-th column is

$$0^{3^m - 3^{m-1} - \cdots - 3^0}(1^{3^m}\bar{1}^{3^m}0^{3^m})^\infty.$$

Since $3^m - 3^{m-1} - \cdots - 3^0 = (3^m+1)/2$, the generating function for the sum of the digits of the balanced ternary representation of n is

$$A(z) = \frac{1}{1-z}\sum_{m\geq 0} z^{(3^m+1)/2}\frac{1-z^{3^m}}{1+z^{3^m}+z^{2\cdot 3^m}} = \frac{1}{1-z}\sum_{m\geq 0} z^{(3^m+1)/2}\frac{(1-z^{3^m})^2}{1-z^{3^{m+1}}}.$$

4 The k-ary Numeration System and Morphisms

The column pattern for k-ary numbers is

$$(\alpha_0^{k^m}\alpha_1^{k^m}\ldots\alpha_{k-1}^{k^m})^\infty. \quad (6)$$

Here the row sum of the n-th row, where

$$n = \sum_{k\geq 0} b_k k^m, \quad \text{is} \quad \sum_{k\geq 0}\alpha_{b_k}k^m.$$

That is, each digit b is "weighted" by α_b. For the digital sum, $\alpha_b = b$.

The special form of (6) implies that there is a "morphism" that underlies the construction; for the digital sum it is $j \to j, j+1, j+2, \ldots, j+k-1$. For example, when $k=3$ we get the sequence $s_3(0), s_3(1), s_3(2), \ldots$ as the limit (i.e., fixed-point) of a morphism noted by Robert G. Wilson in A053735, which gives us successively

$$0 \to 012 \to 012\ 123\ 234 \to 012123234\ 123234345\ 234345456 \to \cdots$$

This limit will exist for any morphism of the form

$$j \to j+\alpha_0, j+\alpha_1, \ldots, j+\alpha_{k-1}, \tag{7}$$

so long as $\alpha_0 = 0$.

Theorem 2. *The row sums of the column pattern (6) are generated by the morphism (7) so long as $\alpha_0 = 0$.*

Proof. The column pattern (6) is invariant under the following two-step operation: (a) Take column m and replace each entry in the column by k identical entries, calling the new column C'_{m+1}. (b) Form a new column C'_0 with the pattern $(01\cdots(k-1))^\infty$. The invariance is that $C'_m = C_m$ for $m = 0, 1, 2, \ldots$.

A row sum j under operations (a) and (b) becomes the k row sums $j+\alpha_0, j+\alpha_1, \ldots, j+\alpha_{k-1}$. This is the morphism (7). □

With the pattern (6) equation (5) gives us the theorem below.

Theorem 3. *If k is an integer with $k \geq 2$ and $\alpha_0 = 0$, then the generating function of the limit of the morphism (7) is*

$$A(z) = \frac{1}{1-z}\sum_{m\geq 0}\frac{\alpha_1 z^{k^m} + \alpha_2 z^{2k^m} + \cdots + \alpha_{k-1} z^{(k-1)z^m}}{1 + z^{k^m} + z^{2k^m} + \cdots + z^{(k-1)k^m}}. \tag{8}$$

Note that Theorem 1 is the special case where $\alpha_i = i$ for $i = 0, 1, \ldots, k-1$. The second equality in Theorem 1 follows from the fact that $z + 2z + \cdots + (k-1)z^{k-1} = (z - kz^k + (k-1)z^{k+1})/(1-z)^2$. We now return to the non-zero count function, $c_k(n)$, which can be expressed without the inner sums used in (8).

Corollary 4. *The generating function of $c_k(n)$ is*

$$C_k(z) = \frac{1}{1-z}\sum_{m\geq 1}\frac{z^{k^{m-1}} - z^{k^m}}{1 - z^{k^m}}. \tag{9}$$

Proof. Here the morphism is $j \to j, j+1, \ldots, j+1$ and so Theorem 3 gives us the first equality below.

$$\begin{aligned} C_k(z) &= \frac{1}{1-z}\sum_{m\geq 0}\frac{z^{k^m} + z^{2k^m} + \cdots + z^{(k-1)k^m}}{1 + z^{k^m} + z^{2k^m} + \cdots + z^{(k-1)k^m}} \\ &= \frac{1}{1-z}\sum_{m\geq 0}\frac{(1-z^{k^{m+1}})/(1-z^{k^m}) - (1-z^{k^m})/(1-z^{k^m})}{(1-z^{k^{m+1}})/(1-z^{k^m})} \end{aligned} \tag{10}$$

Cancelling common denominators and simplifying gives (9). □

Other morphisms would give counts of the number of times individual digits occur in the obvious way. For example, $j \to j, j, j+1, j$ is the morphism for the number of 2's that occur in the 4-ary expansion of n (here the pattern is $(0^{4^m} 0^{4^m} 1^{4^m} 0^{4^m})^\infty$).

Theorem 5. *Let d be an integer with $0 < d < k$. The generating function for the number of digits equal to d in the k-ary expansion of n is*

$$\frac{1}{1-z} \sum_{m \geq 0} \frac{z^{dk^m}}{1 + z^{k^m} + z^{2k^m} + \cdots + z^{(k-1)k^m}} = \frac{1}{1-z} \sum_{m \geq 0} \frac{z^{dk^m}(1 - z^{k^m})}{1 - z^{k^{m+1}}}. \tag{11}$$

The generating function for the number of 0 digits in the k-ary expansion of n is

$$\frac{1}{1-z} \sum_{m \geq 0} \frac{z^{k^{m+1}}}{1 + z^{k^m} + z^{2k^m} + \cdots + z^{(k-1)k^m}} = \frac{1}{1-z} \sum_{m \geq 0} \frac{z^{k^{m+1}}(1 - z^{k^m})}{1 - z^{k^{m+1}}}. \tag{12}$$

Proof. Equation (11) follows from Theorem 3 with the morphism $j \to j, \ldots, j, j+1, j, \ldots, j$ where the $j+1$ occurs in position d, counting from 0. To prove (12) we use the generating function

$$T(z) \;=\; \frac{1}{1-z} \sum_{m \geq 0} z^{k^m}$$

for $1 + \lfloor \log_k n \rfloor$, which is the number of k-ary digits in n. Adding (10) and (12) we clearly obtain $T(z)$.

A second way of finishing the proof is to note that the column pattern

$$0^{k^m} (1^{k^m} 2^{k^m} \cdots (k-1)^{k^m} 0^{k^m})^\infty$$

also describes the k-ary listing of numbers. The useful aspect of expressing it this way is that the leading 0s are correspond to the initial 0^{k^m} above. Thus the pattern for counting (non-leading) 0s is

$$0^{k^m} (0^{k^m} 0^{k^m} \cdots 0^{k^m} 1^{k^m})^\infty.$$

According to (5) the numerator inside the sum of the generating function is $z^{b_m} z^{(u_m - 1)a_m} = z^{k^m} z^{(k-1)k^m} = z^{k^{m+1}}$, as desired. ☐

4.1 Digit counts in specific positions

Let $C(n, r, d)$ be the number of 1 bits in the binary representation of n that are in positions that are congruent to r mod d. As usual, the "positions" are indexed starting at 0 on the right. For example, $888 = (1101111000)_2$, so $C(888, 0, 3) = 3$, $C(888, 1, 3) = 1$ and $C(888, 2, 3) = 2$.

Theorem 6. *For all integers $d \geq 0$ and integers r with $0 \leq r < d$,*

$$\sum_{n \geq 0} C(n, r, d) z^n \;=\; \sum_{m \geq 0} \frac{z^{2^{r+dm}}}{1 + z^{2^{r+dm}}}. \tag{13}$$

OEIS	description	comment	pattern
A023416	0s in binary	same as A080791	$0^{2^m}(0^{2^m}1^{2^m})^\infty$
A000120	1s in binary	digital sum, base 2	$(0^{2^m}1^{2^m})^\infty$
A077267	0s in base 3	same as A081602	$0^{3^m}(0^{3^m}0^{3^m}1^{3^m})^\infty$
A062756	1s in base 3		$(0^{3^m}1^{3^m}0^{3^m})^\infty$
A081603	2s in base 3		$(0^{3^m}0^{3^m}1^{3^m})^\infty$
A160380	0s in base 4		$0^{4^m}(0^{4^m}0^{4^m}0^{4^m}1^{4^m})^\infty$
A160381	1s in base 4		$(0^{4^m}1^{4^m}0^{4^m}0^{4^m})^\infty$
A160382	2s in base 4		$(0^{4^m}0^{4^m}1^{4^m}0^{4^m})^\infty$
A160383	3s in base 4		$(0^{4^m}0^{4^m}0^{4^m}1^{4^m})^\infty$
A160384	non-0 base 3		$(0^{3^m}1^{3^m}1^{3^m})^\infty$
A160385	non-0 base 4		$(0^{4^m}1^{4^m}1^{4^m}1^{4^m})^\infty$
A053735	digital, base 3		$(0^{3^m}1^{3^m}2^{3^m})^\infty$
A053737	digital, base 4		$(0^{4^m}1^{4^m}2^{4^m}3^{4^m})^\infty$
A034968	digital, factorial base	see also A139365	
A065363	digital, balanced base 3		$0^{(3^m+1)/2}(1^{3^m}\bar{1}^{3^m}0^{3^m})^\infty$
A139351	1s or 3s in base 4	1s in even positions in binary	$(0^{4^m}1^{4^m}0^{4^m}1^{4^m})^\infty$
A139352	2s or 3s in base 4	1s in odd positions in binary	$(0^{4^m}0^{4^m}1^{4^m}1^{4^m})^\infty$

Table 1: Relevant sequences in OEIS.

Proof. Let $B(r,d) = \{s \in \mathbb{Z} : 1 \le s < 2^d \text{ and } \lfloor s/2^r \rfloor \text{ is odd}\}$; in other words, the d bit binary numbers with the r-th bit equal to 1. Let $\bar{B}(r,d) = \{0,1,\ldots,2^d-1\} \setminus B(r,d)$. For example, $B(1,3) = \{2,3,6,7\}$ and $B(1,3) = \{0,1,4,5\}$.

Now consider the number n written both in binary and in base 2^d. Note that, in the binary representation of n, the number of 1 bits in positions that are congruent to r mod d is the same as the number of digits from the set $B(r,d)$ in the 2^d-ary representation of n. Thus we may apply Theorem 3 to get the generating function

$$\sum_{n\ge 0} C(n,r,d) z^n \;=\; \sum_{m\ge 0} \frac{\sum_{s\in B(r,d)} z^{s2^{dm}}}{\sum_{0\le s<2^d} z^{s2^{dm}}}$$

Note that the numerator above can be written

$$\sum_{s\in B(r,d)} z^{s2^{dm}} \;=\; z^{2^r 2^{dm}} \sum_{s\in \bar{B}(r,d)} z^{s2^{dm}}$$

and the denominator as

$$\sum_{0\le s<d} z^{s2^{dm}} = \sum_{s\in B(r,d)} z^{s2^{dm}} + \sum_{s\in \bar{B}(r,d)} z^{s2^{dm}} = (1 + z^{2^r 2^{dm}}) \sum_{s\in \bar{B}(r,d)} z^{s2^{dm}}.$$

Canceling the common sum gives the right hand side of (13). $\square$

The following corollary allows us to give a generating function for A139351 and A139352.

Corollary 7. *The generating function for the number of* 1*'s in even positions in the binary expansion of* n*, and the corresponding generating function for the number of* 1*'s in odd positions, are given below.*

$$\frac{1}{1-z}\sum_{m\geq 0}\frac{z^{4^m}}{1+z^{4^m}}, \qquad \frac{1}{1-z}\sum_{m\geq 0}\frac{z^{2\cdot 4^m}}{1+z^{2\cdot 4^m}}.$$

5 Multi-Radix Numeration Systems

In this section we consider numbers written in the multi-radix base $k_0 \times k_1 \times k_2 \times \cdots$. If each $k_i = k$ then we get the k-ary numeration system considered in the previous section. It will prove useful to adopt the following notation: (a) $k'_j = k_j - 1$, (b) $\bar{k}_j = k_0 k_1 \cdots k_{j-1}$, with the usual convention for the empty product, $\bar{k}_0 = 1$. Then the column pattern is

$$(\alpha_0^{\bar{k}_m}\alpha_1^{\bar{k}_m}\cdots\alpha_{k'_m}^{\bar{k}_m})^\infty.$$

Theorem 8. *The generating function for the digital sum of the number* n *written in the multi-radix base* $k_0 \times k_1 \times k_2 \times \cdots$ *is*

$$\frac{1}{1-z}\sum_{m\geq 0}\frac{z^{\bar{k}_m}+2z^{2\bar{k}_m}+\cdots+k'_m z^{k'_m\cdot\bar{k}_m}}{1+z^{\bar{k}_m}+z^{2\bar{k}_m}+\cdots+z^{k'_m\cdot\bar{k}_m}}.$$

Proof. The generating function is (5) with $b_m = 0$, $\alpha_i = i$, $u_m = k'_m$, and $a_m = \bar{k}_m$. □

In the factorial base, $k_j = j+1$, so that $\bar{k}_j = j!$. For example, $99 = 3\cdot 4! + 0\cdot 3! + 2\cdot 2! + 1\cdot 1!$. We now obtain a generating function for A034968 in the following corollary.

Corollary 9. *The generating function for the digital sum of the number* n *written in the factorial base is*

$$\frac{1}{1-z}\sum_{m\geq 1}\frac{z^{m!}+2z^{2m!}+\cdots+mz^{m\cdot m!}}{1+z^{m!}+z^{2m!}+\cdots+z^{m\cdot m!}}.$$

Proof. This follows directly from the previous theorem. Note that the numerator is zero when $m = 0$, so that the summation starts at 1. □

6 Final Remarks

It is also possible to approach the derivation of the generating functions used here using "divide-and-conquer" recurrence relations. See Stephan [8] for examples of this approach in the $k = 2$ case. For example the recurrence relation corresponding to the morphism (7) is $a(0) = 0$ and $a(km + i) = \alpha_i + a(m)$ for integer indices i with $0 \leq i < k$. These recurrence relations are very useful for actually computing the sequences.

As we have shown here, finding a generating function for the sum of the digits is straightforward when dealing with a simple radix, or a mixed radix system where each positional multiplier is a multiple of the previous one. When this does not hold, the problem is much more difficult.

The simplest example of such a system is the Zeckendorf [9] or "base Fibonacci" representation (A014417, digital sum in A007895). Attempting the same sort of one digit at a time approach, the low order digit is the infinite Fibonacci word, A003849. Since this sequence includes arbitrarily long repeated segments, but is not periodic, it does not have a rational generating function.

References

[1] J.-P. Allouche and J. O. Shallit, *Automatic Sequences*, Cambridge University Press, 2003.

[2] J. Bésineau, Indépendance statistique d'ensembles liés à la fonctions "somme des chiffres", *Acta Arith.* **20** (1972), 401–416.

[3] J. Coquet and P. Toffin, Représentations des entiers naturels et indépendance statistique, *Bull. Sci. Math.*, **105** (1981), 289–298.

[4] D. E. Knuth, *The Art of Computer Programming, Volume 2, Seminumerical Algorithms*, 3rd edition, Addison-Wesley, 1998.

[5] D. E. Knuth, *The Art of Computer Programming, Volume 4, Fascicle 1: Bitwise Tricks & Techniques and Binary Decision Diagrams*, Addison-Wesley, 2009.

[6] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, second edition, 1994.

[7] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, www.research.att.com/~njas/sequences/, accessed April 2009.

[8] R. Stephan, Divide-and-conquer generating functions, part I: elementary sequences, http://arxiv.org/abs/math/0307027, 2003.

[9] 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.

2000 *Mathematics Subject Classification*: Primary 05A15; Secondary 05A10.
Keywords: Generating function, numeration system, digital sum, sideways addition, digit counting, morphism, factorial base, balanced ternary numbers.

(Concerned with sequences A000120, A003849, A007895, A014417, A023416, A034968, A053735, A053737, A062756, A065363, A077267, A080791, A081602, A081603, A139351, A139352, A139365, A160380, A160381, A160382, A160383, A160384, and A160385.)

Received June 19 2009; revised version received July 11 2009. Published in *Journal of Integer Sequences*, July 12 2009.

Return to Journal of Integer Sequences home page.