A DIGITAL BINOMIAL THEOREM

HIEU D. NGUYEN

Abstract. We present a triangle of connections between the Sierpinski triangle, the sum-of-digits function, and the Binomial Theorem via a one-parameter family of Sierpinski matrices, which encodes a digital version of the Binomial Theorem.

1. Introduction

It is well known that Sierpinski's triangle can be obtained from Pascal's triangle by evaluating its entries, known as binomial coefficients, mod 2:

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \; 10 \; 10 \; 5 \quad 1 \\
1 \quad 6 \; 15 \; 20 \; 15 \; 6 \quad 1 \\
1 \quad 7 \; 21 \; 35 \; 35 \; 21 \; 7 \quad 1 \\
\cdots \\
\text{Pascal's triangle}
\end{array}
\qquad
\begin{array}{c}
\text{mod } 2 \\
\longrightarrow
\end{array}
\qquad
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 0 \quad 1 \\
1 \quad 1 \quad 1 \quad 1 \\
1 \quad 0 \quad 0 \quad 0 \quad 1 \\
1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \\
1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \\
1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \\
\cdots \\
\text{Sierpinski's triangle}
\end{array}
\tag{1}
$$

Pascla's triangle is, of course, constructed by inserting the binomial coefficient $\binom{n}{k}$ in the k-th position of the n-th row, where the first row and first element in each row correspond to $n = 0$ and $k = 0$, respectively. Binomial coefficients have a distinguished history and appear in the much-celebrated Binomial Theorem:

Theorem 1 (Binomial Theorem)**.**

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}, \qquad n \in \mathbb{N}, \tag{2}$$

where $\binom{n}{k}$ *are defined in terms of factorials:*

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

In this article, we demonstrate how the Binomial Theorem in turn arises from a one-parameter generalization of the Sierpinski triangle. The connection between them is given by the sum-of-digits function, $s(k)$, defined as the sum of the digits in the binary representation of k (see [1]). For example, $s(3) = s(1\cdot 2^1 + 1\cdot 2^0) = 2$. Towards this end, we begin with a well-known matrix formulation of Sierpinski's triangle that demonstrates its fractal nature (see [5], p.246). Define a sequence of matrices S_n of size $2^n \times 2^n$ recursively by

$$S_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{3}$$

and

$$S_{n+1} = S_1 \otimes S_n \tag{4}$$

Date: 12-9-2014.
2010 *Mathematics Subject Classification.* Primary 11.
Key words and phrases. Binomial theorem, Sierpinski triangle, sum-of-digits.

for $n > 1$. Here, the operation $\otimes$ denotes the Kronecker product of two matrices. For example, S_2 and S_3 can be computed as follows:

$$S_2 = S_1 \otimes S_1 = \begin{pmatrix} 1 \cdot S_1 & 0 \cdot S_1 \\ 1 \cdot S_1 & 1 \cdot S_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$S_3 = S_1 \otimes S_2 = \begin{pmatrix} 1 \cdot S_2 & 0 \cdot S_2 \\ 1 \cdot S_2 & 1 \cdot S_2 \end{pmatrix} = \begin{pmatrix} 1 & & & & & & & \\ 1 & 1 & & & & & & \\ 1 & 0 & 1 & & & & & \\ 1 & 1 & 1 & 1 & & & & \\ 1 & 0 & 0 & 0 & 1 & & & \\ 1 & 1 & 0 & 0 & 1 & 1 & & \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Thus, in the limit we obtain Sierpinski's matrix $S = \lim_{n\to\infty} S_n$.

Less well-known is a one-parameter generalization of Sierpinski's triangle in terms of the sum-of-digits function due to Callan [3]. If we define

$$S_1(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \tag{5}$$

and

$$S_{n+1}(x) = S_1(x) \otimes S_n(x) \tag{6}$$

for $n > 1$, then

$$S(x) := \lim_{n\to\infty} S_n(x) = \begin{pmatrix} 1 & & & & & & & & \\ x & 1 & & & & & & & \\ x & 0 & 1 & & & & & & \\ x^2 & x & x & 1 & & & & & \\ x & 0 & 0 & 0 & 1 & & & & \\ x^2 & x & 0 & 0 & x & 1 & & & \\ x^2 & 0 & x & 0 & x & 0 & 1 & & \\ x^3 & x^2 & x^2 & x & x^2 & x & x & 1 & \\ \cdots & & & & & & & & \ddots \end{pmatrix} \tag{7}$$

Observe that $S_n(1) = S_n$ and $S(1) = S$. The matrix $S(x)$ appears in [3] where Callan defines its entries in terms of the sum-of-digits function $s(k)$. In particular, if we denote $S(x) = (s_{j,k})$ and assume the indices j, k to be non-negative with $(j, l) = (0, 0)$ corresponding to the top left-most entry, then the entries $s_{j,k}$ are defined by

$$s_{j,k} = \begin{cases} x^{s(j-k)}, & \text{if } 0 \le k \le j \text{ and } (k, j-k) \text{ is carry-free} \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where the notion of carry-free is defined as follows: call a pair of non-negative integers (a, b) *carry-free* if their sum $a + b$ involves no carries when the addition is performed in binary. For example, the pair $(8, 2)$ is carry-free since $8 + 2 = (1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0) + (1 \cdot 2^1) = 10$ involves no carries in binary.

To see why (8) correctly describes (7), we argue by induction. Clearly, $S_1(x)$ satisfies (8). Next, assume that $S_n(x)$ satisfies (8). It suffices to show that every entry $s_{j,k}$ of $S_{n+1}(x)$ satisfies (8). To prove this, we divide $S_{n+1}(x)$, whose size is $2^{n+1} \times 2^{n+1}$, into four sub-matrices A, B, C, D, each of size $2^n \times 2^n$, based on the recurrence

$$S_{n+1}(x) = \begin{pmatrix} S_n(x) & 0 \\ xS_n(x) & S_n(x) \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where $A = D = S_n(x)$, $B = 0$, and $C = xS_n(x)$. We now consider four cases depending on which sub-matrix the element $s_{j,k}$ belongs to.

Case 1: $0 \leq j, k \leq 2^n - 1$. Then $s_{j,k}$ lies in $A = S_n(x)$ and thus (8) clearly holds.

Case 2: $0 \leq j \leq 2^n - 1$, $2^n \leq k \leq 2^{n+1} - 1$. Then $s_{j,k}$ lies in $B = 0$, which implies $s_{j,k} = 0$, and thus (8) holds since $k \geq j$.

Case 3: $2^n \leq j, k \leq 2^{n+1} - 1$. Then $s_{j,k}$ lies in $D = S_n(x)$. Let

$$j = j_0 2^0 + ... + j_n 2^n,$$
$$k = k_0 2^0 + ... + k_n 2^n$$

denote their binary expansions. Observe that $j_n = k_n = 1$. Define $j' = j - 2^n$ and $k' = k - 2^n$ where we delete the digit j_n from j (resp. k_n from k). Then it is clear that $(k, j-k)$ being carry-free is equivalent to $(k', j'-k')$ being carry-free. Moreover, $s(j-k) = s(j'-k')$. We conclude that

$$s_{j,k} = s_{j',k'} = x^{s(j'-k')} = x^{s(j-k)}$$

satisfies (8).

Case 4: $2^n \leq j \leq 2^{n+1} - 1$, $0 \leq k \leq 2^n - 1$. Then $s_{j,k}$ lies in $C = xS_n(x)$. Define $j' = j - 2^n$ and $k' = k$. Then again $(k, j-k)$ being carry-free is equivalent to $(k', j'-k')$ being carry-free. Also, $s(j-k) = s(2^n + j' - k') = 1 + s(j'-k')$. Hence,

$$s_{j,k} = xs_{j',k'} = x^{1+s(j'-k')} = x^{s(j-k)}$$

satisfies (8) as well. This complete the proof.

Callan also proved in the same paper that $S(x)$ generates a one-parameter group, i.e., it satisfies the following additive property under matrix multiplication:

$$S(x)S(y) = S(x+y) \tag{9}$$

We will see that this property encodes a digital version of the Binomial Theorem. For example, equating the $(3,0)$-entry of $S(x+y)$, i.e. $s_{3,0}$, with the corresponding entry of $S(x)S(y)$ yields the identity

$$(x+y)^{s(3)} = x^{s(3)}y^{s(0)} + x^{s(2)}y^{s(1)} + x^{s(1)}y^{s(2)} + x^{s(0)}y^{s(3)}, \tag{10}$$

which simplifies to the Binomial Theorem for $n = 2$:

$$(x+y)^2 = x^2 + 2xy + y^2. \tag{11}$$

The identities corresponding to the $(5,0)$ and $(7,0)$-entries of $S(x+y)$ are

$$(x+y)^{s(5)} = x^{s(5)}y^{s(0)} + x^{s(4)}y^{s(1)} + x^{s(1)}y^{s(4)} + x^{s(5)}y^{s(0)} \tag{12}$$

and

$$\begin{aligned}(x+y)^{s(7)} = {} & x^{s(7)}y^{s(0)} + x^{s(6)}y^{s(1)} + x^{s(5)}y^{s(2)} + x^{s(4)}y^{s(3)} \\ & + x^{s(3)}y^{s(4)} + x^{s(2)}y^{s(5)} + x^{s(1)}y^{s(6)} + x^{s(2)}y^{s(1)},\end{aligned} \tag{13}$$

respectively. Observe that (12) is equivalent to (10) while (13) simplifies to the Binomial Theorem for $n = 3$.

More generally, property (9) can be restated as a digital version of the Binomial Theorem:

Theorem 2 (Digital Binomial Theorem)**.** *Let $m \in \mathbb{N}$. Then*

$$(x+y)^{s(m)} = \sum_{\substack{0 \leq k \leq m \\ (k, m-k) \text{ carry-free}}} x^{s(k)}y^{s(m-k)}. \tag{14}$$

We note that (26) appears implicitly in Callan's proof of (9). The rest of this article is devoted to proving Theorem 2 independently of (9) and demonstrating that it is equivalent to the Binomial Theorem when $m = 2^n - 1$.

2. Proof of the Digital Binomial Theorem

There are many known proofs of the Binomial Theorem. The standard combinatorial proof relies on enumerating n-element permutations that contain the symbols x and y and then counting those permutations that contain k copies of x. For example, the expansion

$$(x+y)^2 = xx + xy + yx + yy \tag{15}$$

gives all 2-element permutations that contain x and y. Then the number of permutations that contain k copies of x is given by $\binom{2}{k}$. Thus, (15) corresponds to (2) with $n = 2$:

$$(x+y)^2 = \binom{2}{0}x^2 + \binom{2}{1}xy + \binom{2}{2}y^2. \tag{16}$$

To establish that (16) is equivalent to (10), we consider the following *digital* binomial expansion: given two sets of digits, $S_0 = \{x_0, y_0\}$ and $S_1 = \{x_1, y_1\}$, we can represent all ways of constructing a 2-digit number z_0z_1, where $z_0 \in S_0$ and $z_1 \in S_1$, by the expansion

$$(x_0 + y_0)(x_1 + y_1) = x_0x_1 + x_0y_1 + y_0x_1 + y_0y_1, \tag{17}$$

which we rewrite as

$$(x_0 + y_0)(x_1 + y_1) = (x_0^1x_1^1)(y_0^0y_1^0) + (x_0^1x_1^0)(y_0^0y_1^1) + (x_0^0x_1^1)(y_0^1y_1^0) + (x_0^0x_1^0))y_0^1y_1^1). \tag{18}$$

If we now assume that $x_0 = x_1 = x$ and $y_0 = y_1 = y$, then each term on the right-hand side of (18) has the form

$$x_0^{d_0}x_1^{d_1}y_0^{1-d_0}y_1^{1-d_1} = x^{s(k)}y^{s(3-k)},$$

where $k = d_02^0 + d_12^1$ and $3 - k = (1-d_0)2^0 + (1-d_1)2^1$. It follows that (18) reduces to (10). On the other hand, (17) reduces to (15). Thus, we have shown that Theorem 2 for $m = 3$ is equivalent to the Binomial Theorem for $n = 2$.

To extend the proof to integers of the form $m = 2^n - 1$, we consider n sets of digits, $S_k = \{x_k, y_k\}$, where $k = 0, 1, \ldots, n-1$. The expansion

$$\prod_{k=0}^{n-1}(x_k + y_k) = \sum_{\substack{z_k \in S_k \\ \forall k=0,1,\ldots,n-1}} z_0 \ldots z_{n-1} = \sum_{\substack{d_k \in \{0,1\} \\ \forall k=0,1,\ldots,n-1}} x_0^{d_0} \cdots x_{n-1}^{d_{n-1}} y_0^{1-d_0} \cdots y_{n-1}^{1-d_{n-1}} \tag{19}$$

represents all ways of constructing an n-digit number $z = z_0z_1 \ldots z_{n-1}$ with $z_k \in S_k$ for $k = 0, 1, \ldots, n-1$. Then substituting $x_k = x$ and $y_k = y$ for all such k into (19) yields

$$(x+y)^n = \sum_{d_0,\ldots,d_{n-1} \in \{0,1\}} x^{d_0+\ldots+d_{n-1}} y^{n-(d_0+\ldots+d_{n-1})}, \tag{20}$$

or equivalently,

$$(x+y)^{s(2^n-1)} = \sum_{k=0}^{2^n-1} x^{s(k)}y^{s(2^n-1-k)}, \tag{21}$$

where if we define $k = d_02^0 + \ldots + d_{n-1}2^{n-1}$, then $s(k) = d_0 + \ldots + d_{n-1}$ and

$$s(2^n - 1 - k) = s(2^n - 1) - s(k) = n - (d_0 + \ldots + d_{n-1}).$$

Moreover, k ranges from 0 to $2^n - 1$ since $d_0, \ldots, d_{n-1} \in \{0, 1\}$. This justifies Theorem 1. On the other hand, given k between 0 and n, the number of permutations $(d_0, \ldots, d_{n-1})$ containing k 1's is equal to $\binom{n}{k}$. Thus, (20) reduces to (2). This proves that Theorem 1 is equivalent to the Binomial Theorem.

To complete the proof of Theorem 1 for any non-negative integer m, we first expand m in binary:

$$m = m_{i_0}2^{i_0} + \ldots + m_{i_{n-1}}2^{i_{n-1}},$$

where we only record its 1's digits so that $m_{i_k} = 1$ for all $k = 0, \ldots, n-1$. Then $s(m) = m_{i_0} + \cdots + m_{i_{n-1}} = n$. Just as before, we use the expansion (19) to derive (20), but this time we rewrite (20) as

$$(x+y)^{s(m)} = \sum_{\substack{0 \le k \le m \\ (k,m-k) \text{ carry-free}}} x^{s(k)}y^{s(m-k)}, \tag{22}$$

where we define

$$k = d_0 2^{i_0} + \ldots + d_{n-1} 2^{i_{n-1}}. \tag{23}$$

Then $s(k) = d_0 + \ldots + d_{n-1}$ and since $m_{i_k} = 1$ for all $k = 0, \ldots, n-1$, we have

$$\begin{aligned} m - k &= (m_{i_0} 2^{i_0} + \ldots + m_{i_{n-1}} 2^{i_{n-1}}) - (d_0 2^{i_0} + \ldots + d_{n-1} 2^{i_{n-1}})) \\ &= (1 - d_0) 2^{i_0} + \ldots + (1 - d_{n-1}) 2^{i_{n-1}}. \end{aligned}$$

It follows that

$$\begin{aligned} s(m - k) &= (1 - d_0) + \ldots + (1 - d_{n-1}) \\ &= n - (d_0 + \ldots + d_{n-1}). \end{aligned}$$

Moreover, it is clear that $0 \le k \le m$ and $(k, m-k)$ is carry-free. Conversely, every non-negative integer k with $(k, m-k)$ carry-free must have representation in the form (23); otherwise, the sum $k + (m-k)$ requires a carry in any non-zero digit of k where the corresponding digit of m in the same position is zero. Thus, Theorem 1 holds for any non-negative integer m.

To complete our story we explain why Sierpinski's triangle appears in the reduction of Pascal's triangle's mod 2 by relating binomial coefficients with the sum-of-digits function. Define the carry function $c(n, k)$ to be the number of carries needed to add k and $n - k$ in binary. A theorem of Kummer's (see [4]) tells us that the p-adic valuation of binomial coefficients is given by the carry function.

Theorem 3 (Kummer). *Let p be a prime integer. Then the largest power of p that divides $\binom{n}{k}$ equals $c(n, k)$.*

Kummer's theorem now explains the location of 0's and 1's in Sierpinski's triangle, assuming that its entries are defined by

$$s_{n,k} = \binom{n}{k} \mod 2. \tag{24}$$

Let $p = 2$. If $(k, n-k)$ is carry free, then $c(n,k)$=0 and therefore the largest power of 2 dividing $\binom{n}{k}$ is $2^0 = 1$. In other words, $\binom{n}{k}$ is odd and hence, $s_{n,k} = 0$. On the other hand, if $(k, n-k)$ is *not* carry-free, then $c(n, k) \ge 1$ and so the largest power of 2 dividing $\binom{n}{k}$ is at least 1. Therefore, $\binom{n}{k}$ is even and hence, $s_{j,k} = 0$. This proves that definition (24) for Sierpinski's triangle is equivalent to definition (8) in terms of carry-free pairs with $x = 1$.

Lastly, it is known that the failure of the sum-of-digits function to be additive is characterized by the carry function. In particular, we have (see [2])

$$s(k) + s(n-k) - s(n) = c(n, k) \tag{25}$$

It follows that $(k, n-k)$ is carry-free if and only if $s(k) + s(n-k) = s(n)$. Thus, it is fitting that the Digital Binomial Theorem can be restated purely in terms of the additivity of the sum-of-digits function:

$$(x + y)^{s(m)} = \sum_{\substack{0 \le k \le m \\ s(k)+s(m-k)=s(m)}} x^{s(k)} y^{s(m-k)}. \tag{26}$$

References

[1] J.-P. Allouche and J. Shallit, *The Ubiquitous Prouhet-Thue-Morse Sequence*, Sequences and Their applications, Proc. SETA'98 (Ed. C. Ding, T. Helleseth, and H. Niederreiter). New York: Springer-Verlag, pp. 1–16, 1999.

[2] T. Ball, T. Edgar, and D. Juda, *Dominance Orders, Generalized Binomial Coefficients, and Kummer's Theorem*, Mathematics Magazine **87** (2014), No. 2, 135-143.

[3] D. Callan, *Sierpinski's Triangle and the Prouhet-Thue-Morse Word*, 2006. arXiv:math/0610932.

[4] Andrew Granville, *Arithmetic properties of binomial coefficients I: Binomial coefficients modulo prime powers*, in Organic mathematics, CMS Conf. Proc. vol. 20, Amer. Math. Soc., 1997, 253-276.

[5] Willi-Hans Steeb, The Nonlinear Workbook: 5th Edition, World Scientific Publishing Company, 2011.

Department of Mathematics, Rowan University, Glassboro, NJ 08028.
E-mail address: nguyen@rowan.edu