Chapter 8

Pascal’s Triangle: Cellular Automata and Attractors

Mathematics is often defined as the science of space and number [. . .] It was not until the recent resonance of computers and mathematics that a more apt definition became fully evident: mathematics is the science of patterns.

Lynn Arthur Steen, 1988

Being introduced to the Pascal triangle for the first time, one might think that this mathematical object was a rather innocent one. Surprisingly it has attracted the attention of innumerable scientists and amateur scientists over many centuries. One of the earliest mentions (long before Pascal’s name became associated with it) is in a Chinese document from around 1303.1 Boris A. Bondarenko,2 in his beautiful recently published book, counts several hundred publications which have been devoted to the Pascal triangle and related problems just over the last two hundred years. Prominent mathematicians as well as popular science writers such as Ian Stewart,3 Evgeni B. Dynkin and Wladimir A. Uspenski,4 and Stephen Wolfram5 have devoted articles to the marvelous relationship between elementary number theory and the geometrical patterns found in the Pascal triangle. In chapter 2 we introduced one example: the relation between the Pascal triangle and the Sierpinski gasket.

This relationship is indeed a wonderful marvel, and we want to take this opportunity to demonstrate how approaching one mathematical question from totally different angles can beautifully lead to a thorough understanding of

---

1See figure 2.24 in chapter 2.
Capturing Pascal’s Triangle

Three approaches to the patterns in Pascal’s triangle.

Figure 8.1

that matter. Let us restate the problem.\(^6\) Look at the Pascal’s triangle in figure 8.1. It has long been observed that by coloring

- all odd entries black, and
- all even entries white,

we obtain a geometrical pattern which is very closely related to the Sierpinski gasket. Figure 8.1 shows the first 5 rows and the beginning of this pattern formation (first, the black cells only outline a triangle) and figure 8.2 shows the first 128 rows.\(^7\) In fact, the more rows we take into account (e.g., 256, 512, etc.), the more details of the Sierpinski gasket become visible in the geometric pattern.

But we can also use different coloring rules. This leads to all kinds of amazing fractal structures in the triangle. Thus, it is a very interesting question whether there is a way to describe these global pattern formations and how we can find their mathematical foundations for them.

The most important mathematical interpretation of Pascal’s triangle is through binomial coefficients, i.e., the coefficients of the polynomials:

\[
\begin{align*}
(1 + x)^0 &= 1 \\
(1 + x)^1 &= 1 + 1x \\
(1 + x)^2 &= 1 + 2x + 1x^2 \\
&
\vdots \\
(1 + x)^n &= a_0 + a_1 x + \cdots + a_n x^n.
\end{align*}
\]


\(^7\)See also figure 2.26 in chapter 2.
These coefficients\(^8\) are explicitly given by

\[
a_k = \binom{n}{k} = \frac{n!}{(n-k)!k!}, \quad 0 \leq k \leq n
\]  
\[(8.1)\]

where, as usual, factorial \(n\) is defined as

\[n! = 1 \cdot 2 \cdot 3 \cdots n\]

for \(n \geq 1\), and \(0! = 1\).\(^9\) Here are some particular cases, which directly follow from these definitions.

\[
\begin{align*}
\binom{n}{0} &= 1, & \binom{n}{1} &= n, & \binom{n}{n-1} &= n, & \binom{n}{n} &= 1.
\end{align*}
\]

Moreover,

\[
\binom{n}{k} = \binom{n}{n-k}.
\]

In other words, introducing a coordinate system for the cells in the triangular array as in in figure 8.1, where \(n = 0, 1, 2, \ldots\) is the row index and \(k = 0, 1, 2, \ldots\) is the column index, then the entry in cell with coordinates \((n, k)\) is \(\binom{n}{k}\).

---

\(^8\)The notation \(\binom{n}{k}\) was introduced by Andreas von Ettingshausen in his book *Die kombinatorische Analysis*, Vienna, 1826.

\(^9\)For later consideration in the context of cellular automata, we also adopt the convention \(\binom{n}{k} = 0\) for \(k < 0\) and \(k > n\).
Thus, one approach to the patterns in Pascal’s triangle would be to understand the divisibility properties of binomial coefficients. However, computing the entries $a_{k}^{n}$ according to eqn. (8.1) for figures like 8.2 does not lead very far. The reason is that factorials grow extremely rapidly.

The number $100!$ has 158 digits,

$$100! = 9332621544394415268169923885626670049071$$

$$5968264381621468592963895217599993229915$$

$$6089414639761565182862536979208272237582$$

$$5118521091686400000000000000000000,$$

and $1000!$ about 2568 digits, which surely is beyond the range of common computer arithmetic.\(^{10}\)

As a first step to overcome these difficulties we use the recursive definition of Pascal’s triangle (as indicated in figure 8.1) which is obtained from the addition rule\(^{11}\)

\[\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.\] (8.2)

This fundamental relation avoids the computation of large factorials. However, the binomial coefficients themselves also grow rapidly as the row index $n$ increases. Already in row $n = 34$ we find an entry

\[\binom{34}{17} = 2,333,606,220 > 2^{34} - 1 = 2,147,483,647\]

which cannot be represented exactly in normal computer arithmetic. Fortunately, we do not need the actual numerical values of the binomial coefficients when testing for divisibility. For example, the decision whether a binomial coefficient is odd or even follows directly from the addition rule. Observe that $\binom{n+1}{k}$ is odd provided $\binom{n}{k-1}$ is odd and $\binom{n}{k}$ is even, or vice versa. Systematically we have:

\(^{10}\)The estimate of 2568 digits is obtained by a famous formula developed by James Stirling in 1730 which approximates $n! \approx \sqrt{2\pi n}(n/e)^n$, where $e = 2.71828\ldots$ denotes Euler’s number.

\(^{11}\)See section 2.3.
This elementary observation is not only of computational importance; it also provides the link to cellular automata. This is another approach to Pascal’s triangle which will be explored in the following. We will see that there is a whole class of cellular automata which are closely related to the evolution of divisibility patterns in Pascal’s triangle.

However, running a cellular automaton and testing divisibility properties of binomial coefficients have a common property, namely, that they are local (or microscopic) procedures. They allow the generation of a geometric pattern but do not at all explain the global (or macroscopic) appearance of the pattern. For example, why do we begin to see the Sierpinski gasket when coloring the odd entries in Pascal’s triangle?

To address this problem we once again will bring iterated function systems (IFS) into play. If you recall section 5.4, this does not come as a total surprise and you might have a initial vague idea how this approach to Pascal’s triangle could look. We will guide you to this point and explore its relation to divisibility properties and cellular automata, and you will watch the pieces of the puzzles falling into place.

<table>
<thead>
<tr>
<th>(\binom{n}{k-1})</th>
<th>(\binom{n}{k})</th>
<th>(\binom{n+1}{k})</th>
</tr>
</thead>
<tbody>
<tr>
<td>even even even</td>
<td></td>
<td></td>
</tr>
<tr>
<td>odd even odd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>even odd odd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>odd odd even</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8.1 Cellular Automata

Cellular automata have starting points far back in the sciences. In some sense we might say that Pascal’s triangle is the first cellular automaton. Their recent development is rooted in the work of Konrad Zuse, Stanislaw Ulam and John von Neumann and is closely related to the first computing machines. During the 1970’s and 1980’s cellular automata had a strong revival through the work of Stephen Wolfram, who published an interesting survey. Today cellular automata have become a very important modeling and simulation tool in science and technology, from physics, chemistry, and biology, to computational fluid dynamics in airplane and ship design, and to philosophy and sociology.

One-Dimensional Two-State Automaton

The first steps of a one-dimensional cellular automaton with two states (black and white cells).

Cellular automata are perfect feedback machines. More precisely, they are mathematical finite state machines which change the state of their cells step by step. Each cell has one out of $p$ possible states represented by the numbers 0, 1, …, $p - 1$. Sometimes we speak of a $p$-state cellular automaton. The automaton can be one-dimensional where its cells are simply lined up like a chain or two-dimensional where cells are arranged in an array covering the plane. Figure 8.3 shows the first steps of a one-dimensional two-state automaton. Sometimes we like to draw the succeeding steps of a one-dimensional cellular automaton one below the other and call the steps layers. When running the machine it grows layer by layer as shown in figure 8.4.

To run a cellular automaton we need two entities of information: an initial state of its cells (i.e., an initial layer) and a set of rules or laws. These rules describe how the state of a cell in a new layer (in the next step) is determined from the states of a group of cells from the preceding layer. The rules should not depend on the position of the group within the layer. Thus, it can be specified by a look-up table or if possible by a formula. Figure 8.6 shows look-up tables for two-state cellular automata which are given by configurations as in (a) and (b) of figure 8.5. These are just two examples of rules for one-dimensional cellular automata. The look-up table (a) was used in figures 8.3 and 8.4.

---

13 In fact, the automaton can have any dimension $m$, where $m$ is a natural number.
8.1 Cellular Automata

Growing of Layers of a One-Dimensional Automaton

The iterations of the one-dimensional cellular automaton from figure 8.3 are continued. The first 10 steps of are drawn from top to bottom.

![Growing of Layers of a One-Dimensional Automaton](image)

Automata Rules

There are several ways a rule may determine the state of a cell in the succeeding layers. In (a) the state of a new cell is determined by the states of two cells, in (b) by the state of three cells. In (c) and (d) the states of five cells determine the state of a new cell, but note that the position of the new cell with respect to the group is different in (c) and (d).

![Automata Rules](image)

Game of Life

Particular two-dimensional cellular automata became very popular as the *Game of Life* through the work of John Horton Conway in the 1970’s. In the Game of Life each cell is either dead (0) or alive (1) and changes its state according to the states in its immediate neighborhood, including its own state. More precisely, a cell that is alive (symbolized as a black cell) at one time step will stay alive in the next step when precisely two or three cells among its eight neighbors (see figure 8.7) in a square lattice are alive. If more than three neighbors are alive, the cell will die from overcrowdedness. If fewer than two neighbors are alive, the cell will die from loneliness. A dead cell will come to life when surrounded by exactly three live neighbors. Figure 8.8 shows the evolution of the Game of Life in some steps. One of the challenges of the game is to design cell clusters which exhibit a particularly interesting behavior. For example, there are clusters, called *blinkers*, which reproduce...
Automata Look-Up Tables

Two examples of look-up tables. (a) four rules for a configuration based on two cells and two states. (b) eight rules for a configuration based on three cells and two states.

Figure 8.6

Neighborhood in Two-Dimensional Automata

In the Game of Life the neighborhood of a cell decides over life or death. Cell (a) and (b) will stay alive but cell (c) and (d) will die. At (f) a cell will come into life but not at (e).

Figure 8.7

themselves after some steps, gliders move in a certain direction, star ships leave a trace of blinkers, and guns periodically eject gliders.

The rules of the Game of Life are only one choice out of many imaginable sets of rules. For the two possible states and a neighborhood of eight cells generating a new center cell there are \(2^8 \approx 10^{154}\) different possible sets.

Let us briefly touch some variants of the Game of Life. The one-out-of-eight rule is given by the following set of rules: a cell becomes alive if exactly one of its neighbors is alive; otherwise it remains unchanged. Figure 8.9 shows the resulting pattern which evolves after 29 steps starting with just one living cell. Apparently some self-similarity is built into the formation of this pattern.

Another example, the majority rule, is obtained by these conventions: if five or more of the neighborhood of nine cells (including the cell itself) are

The Number of Games
The Game of Life

Six successive steps of the Game of Life. Dots indicate the position of living cells of the previous step. Observe that some of the cell clusters shown exhibit a periodic behavior. The center right one is a so-called glider which slowly moves to the left as long as it does not hit another cell cluster. It takes 4 steps to move one cell to the left.

alive, then this cell will also become or remain alive. Otherwise it will die or remain dead. In other words, the center cell adjusts to the majority in the neighborhood. The resulting patterns resemble some phenomena in statistical physics such as percolation\textsuperscript{14} or Ising spin systems. Figure 8.11 shows some

\textsuperscript{14}See section 9.2 in chapter 9.
One-Out-of-Eight Rule

Starting with just one cell this self-similar pattern evolves after 29 steps.

![Image](image1.png)

Figure 8.9

Experiments which evolve as stable pattern after some 30 steps starting in each case with a random initial distribution of living cells.

NWSE Neighbors

The north, west, south, and east neighbors.

![Image](image2.png)

Figure 8.10

Finally, we consider rules which only take four neighbors (again in a two-dimensional square lattice) into account (see figure 8.10). Following Tommaso Toffoli and Norman Margolus\(^{15}\) we label the center cell by \(C\) = center, and the four neighbors are labelled \(E\) = east, \(W\) = west, \(S\) = south, and \(N\) = north. If we allow two states for each cell of this configuration of five cells (CSWNE) then the state of CSWNE will be given by five binary digits. For example.

Figure 8.11: Two examples for a game with the majority rule. Starting from a random distribution of black cells the game settles down (i.e., further iterations do not change the state of cells) to the patterns shown. For the two images, two different initial distributions were used.

$CSWNE = 11010$ indicates that cells $C$, $S$, and $N$ are alive, while the other two are dead. A complete set of rules can be given by a table of the 32 possible states of $CSWNE$ and the subsequent values of the center cell $C$. Note that for such a configuration there are $2^{32} \approx 4 \cdot 10^9$, i.e., 4 billion possible different tables.

<table>
<thead>
<tr>
<th>$CSWNE$</th>
<th>$C$</th>
<th>$CSWNE$</th>
<th>$C$</th>
<th>$CSWNE$</th>
<th>$C$</th>
<th>$CSWNE$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000</td>
<td>0</td>
<td>01000</td>
<td>1</td>
<td>10000</td>
<td>1</td>
<td>11000</td>
<td>1</td>
</tr>
<tr>
<td>00001</td>
<td>0</td>
<td>01001</td>
<td>1</td>
<td>10001</td>
<td>1</td>
<td>11001</td>
<td>1</td>
</tr>
<tr>
<td>00010</td>
<td>0</td>
<td>01010</td>
<td>1</td>
<td>10010</td>
<td>1</td>
<td>11010</td>
<td>1</td>
</tr>
<tr>
<td>00011</td>
<td>0</td>
<td>01011</td>
<td>1</td>
<td>10011</td>
<td>1</td>
<td>11011</td>
<td>1</td>
</tr>
<tr>
<td>00100</td>
<td>1</td>
<td>01100</td>
<td>0</td>
<td>10100</td>
<td>1</td>
<td>11100</td>
<td>1</td>
</tr>
<tr>
<td>00101</td>
<td>1</td>
<td>01101</td>
<td>0</td>
<td>10101</td>
<td>1</td>
<td>11101</td>
<td>1</td>
</tr>
<tr>
<td>00110</td>
<td>1</td>
<td>01110</td>
<td>0</td>
<td>10110</td>
<td>1</td>
<td>11110</td>
<td>1</td>
</tr>
<tr>
<td>00111</td>
<td>1</td>
<td>01111</td>
<td>0</td>
<td>10111</td>
<td>1</td>
<td>11111</td>
<td>1</td>
</tr>
</tbody>
</table>

Applying the rules from the above table we obtain a familiar pattern. Figure 8.12 shows the 5th and 15th step starting with just one live cell near the lower left corner. Studying the table you can find a rather simple rule for producing the entries. Note that if the center cell $C$ is dead (0) then the new value depends only on cells $W$ and $S$. On the other hand, if the cell is alive (1) then it will remain alive in the next step. In fact, we have just seen an example of how we can construct a two-dimensional automaton with the behavior of
a one-dimensional one. In other words, cells grow layer by layer like the layers of a one-dimensional cellular automaton (although in this example the layers are diagonals and the pattern grows from bottom left to top right).

\^16In fact, the same basic idea works for any given one-dimensional automaton.
8.1 Cellular Automata

Sierpinski Automaton

The first 16 layers of a cellular automaton with look-up table displayed in the upper right.

Many interesting rules can be expressed by a simple formula. For example, the parity rule is given simply by

$$C_{\text{new}} = C_{\text{old}} + S_{\text{old}} + W_{\text{old}} + N_{\text{old}} + E_{\text{old}} \mod 2.$$  \hfill (8.3)

which means that $C_{\text{new}}$ is 0 or 1 if the sum on the right-hand side is even or odd, respectively. Here $E, W, S, N$ and $C$ represent the old and new cell states as indicated by the indices ‘old’ and ‘new’. Thus for $CSWNE = 11010$ we obtain $C_{\text{new}} = 1$, for $CSWNE = 11011$ $C_{\text{new}} = 0$, and so on. Figure 8.13 shows the evolution of this cellular automaton after 13 and 27 steps starting with a square block of $8 \times 8$ black cells.

Pascal’s Triangle...

Let us return to one-dimensional automata. The look-up table used in figure 8.14 reflects the addition of even and odd binomial coefficients. An odd entry is colored black. That is, the evolution of the corresponding cellular automaton will produce the pattern which is obtained from the Pascal triangle when we color cells with odd entries black and cells with even entries white and start with an appropriate initial layer. This is seen in the figure, where we follow the evolution of the first 16 layers starting with the initial layer, which has one black cell.

...and Polynomials

Let us explore the connection between cellular automata and coefficients of polynomials a bit further. First we look at an example involving the powers
of the polynomial \( r(x) = 1 + x \):

\[
\begin{align*}
(r(x))^0 &= 1 \\
(r(x))^1 &= 1 + x \\
(r(x))^2 &= 1 + 2x + x^2 \\
(r(x))^3 &= 1 + 3x + 3x^2 + x^3 \\
(r(x))^4 &= 1 + 4x + 6x^2 + 4x^3 + x^4 \\
&\vdots \\
(r(x))^n &= a_0(n) + a_1(n)x + a_2(n)x^2 + \ldots + a_n(n)x^n.
\end{align*}
\]

Now let \( a_k(n) \) for all integers \( k \) and \( n \geq 0 \) denote the state of cell number \( k \) of the \( n \)-th layer of a one-dimensional automaton.\(^\text{17}\) Starting with \( a_0(0) = 1 \) and \( a_k(0) = 0 \) for \( k \neq 0 \) the rule

\[
a_k(n) = a_{k-1}(n-1) + a_k(n-1)
\]

(8.4)

generates the coefficients of \( (r(x))^n \). Equation (8.4) is nothing else but the addition rule in eqn. (8.2) for binomial coefficients.

Now we want to look at the divisibility properties of \( a_k(n) \) with respect to an integer \( p \). We write

\[
a \equiv b \pmod{p}
\]

provided \( a - b \) is a multiple of \( p \).\(^\text{18}\) Using this language, our test for odd and even binominal coefficients \( \binom{n}{k} \) or cells \( a_k(n) \) is simply to check whether

\[
a_k(n) \equiv 0 \pmod{2} \text{ or } a_k(n) \equiv 1 \pmod{2}.
\]

Moreover, our addition rules (8.2) and (8.4) imply in mod 2 arithmetic

<table>
<thead>
<tr>
<th>(a_{k-1}(n))</th>
<th>(a_k(n))</th>
<th>(a_k(n+1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

This is just the look-up table of the 2-state automaton shown in figure 8.14 where black corresponds to 1 and white to 0. Thus, this figure shows the coefficients of the powers of \( r(x) = 1 + x \) modulo 2.

\(^{17}\)Strictly speaking, this is not a finite state automaton because the numbers \( a_k(n) \) grow beyond all bounds. However, we will arrive at a finite state machine when restricting our attention to the divisibility properties.

\(^{18}\)In other words, \( a \equiv b \pmod{p} \) provided \( a \) and \( b \) differ by a multiple of \( p \). For \( p = 2 \) this means that \( a - b \) is even. Furthermore, \( a \equiv 0 \pmod{2} \) means that \( a \) is even and \( a \equiv 1 \pmod{2} \) means that \( a \) is odd. The notation was introduced by Carl Friedrich Gauss.
8.1 Cellular Automata

Mod 3 Automaton

Cellular automaton generated by the polynomial \( r(x) = 1 + x + x^2 \) and \( p = 3 \). There are two types of indices, a row index \( n \) which runs \( n = 0, 1, 2, \ldots \), and a column index \( k \) which runs in the integers \( \ldots, -2, -1, 0, 1, 2, \ldots \). The initial layer, \( n = 0 \), consists of cells of state 0 except for the cell at \( k = 0 \), which is a cell with state 1. The rule for the cells of the new layer is: \( a_{n+1,k} = a_{n,k-2} + a_{n,k-1} + a_{n,k} \mod 3 \)

Figure 8.15

Generalizations

Now we can generalize in two ways: we can look at coefficients modulo integers other than 2 and we can look at arbitrary polynomials. Let us take the example: \( r(x) = 1 + x + x^2 \):

\[
egin{align*}
(r(x))^0 &= 1 \\
(r(x))^1 &= 1 + x + x^2 \\
(r(x))^2 &= 1 + 2x + 3x^2 + 2x^3 + x^4 \\
(r(x))^3 &= 1 + 3x + 6x^2 + 7x^3 + 6x^4 + 3x^5 + x^6 \\
&
\vdots \\
(r(x))^n &= a_0(n) + a_1(n)x + a_2(n)x^2 + \cdots + a_{2n}(n)x^{2n}.
\end{align*}
\]

Do you see an extension of the addition rule for binomial coefficients, eqn. (8.4)? You can check in the first few lines that the law

\[
a_k(n) = a_{k-2}(n-1) + a_{k-1}(n-1) + a_k(n-1)
\]

holds. A proof of this relation would proceed by induction. When looking at the divisibility properties with respect to \( p = 3 \) we obtain the coefficients

\[
egin{align*}
(r(x))^0 &\rightarrow 1 \\
(r(x))^1 &\rightarrow 1 1 1 \\
(r(x))^2 &\rightarrow 1 2 0 2 1 \\
(r(x))^3 &\rightarrow 1 0 0 1 0 0 1
\end{align*}
\]

and so on. Figure 8.15 shows the evolution of the corresponding 3-state automaton.

Linear Cellular Automata

In a similar way we could start with any polynomial \( r(x) = a_0 + a_1 x + \cdots + a_d x^d \) of degree \( d \) and integer coefficients \( a_i \), and then look at the coefficients of \( (r(x))^n \) modulo some positive integer \( p \) for \( n = 0, 1, 2, \ldots \) and the result would be that the \( k \)-th coefficient of \( (r(x))^{n+1} \) is obtained by an addition
formula involving \( d + 1 \) coefficients from \((r(x))^n\). In other words, given a polynomial with integer coefficients and a positive integer \( p \) there is an associated cellular automaton which generates the coefficients modulo \( p \) of the powers \((r(x))^n, n = 0, 1, 2, \ldots\). Since the look-up table is generated by an addition formula these automata are called \textit{linear cellular automata} (LCA).

The choice of the positive integer \( p \) determines the number of states of the automaton. If \( p = 2 \), i.e., we are considering arithmetic modulo 2, then we have an automaton which can be graphically represented in black and white. For \( p > 2 \) we would need colors to adequately represent the evolution of an automaton. We can often simplify a \( p \)-state to a 2-state automaton by the following modification:

- Cells representing a nonzero coefficient are colored black.
- Cells representing a zero coefficient are colored white.

With this background of linear cellular automata, we can state a number of very interesting problems:

- **Pattern Formation.** Given a polynomial with integer coefficients and a positive integer \( p \), discuss the global pattern formation which evolves when the automaton has produced for a long time.
- **Colors.** What is the relationship between the global patterns which are obtained for different choices of \( p \) and a fixed, given polynomial?
- **Fractal Dimension.** What is the fractal dimension of the global pattern?
- **Higher Dimensions.** Polynomials in one variable generate one-dimensional linear cellular automata. A polynomial in \( m \) variables determines a linear cellular automaton in \( m \) dimensions. How can we generalize the results to \( m \)-dimensional automata?
- **Factorization.** If a polynomial \( r(x) \) is the product of two polynomials \( s(x) \) and \( t(x) \), how is the pattern determined by \( r(x) \) related to the patterns determined by \( s(x) \) and \( t(x) \), and how are the dimensions related?

Actually, the last problem critically depends on the choice of \( p \), the number of states, because what actually counts is whether

\[
r(x) \equiv s(x)t(x) \pmod{p}.
\]

For example, the polynomial \( r(x) = 1 + x \) is irreducible with respect to the integers, i.e., if \( r(x) = s(x)t(x) \), then the factorization must be trivial, i.e., \( s(x) = 1 \) and \( t(x) = 1 + x \). If we use arithmetic modulo \( p \), and \( p \) is not a prime number, however, then \( r(x) = 1 + x \) admits nontrivial factorizations like, for example,

\[
1 + x \equiv (1 + 3x)(1 + 4x) \pmod{6}.
\]

Several of these problems are still wide open while others have been understood only recently through new tools provided by fractal geometry (namely, hierarchical iterated function systems) which stresses again that fractals are more than pretty images.
8.2 Binomial Coefficients and Divisibility

In the remaining part of this chapter we will discuss some of the problems listed at the end of the last section for the particular choice \( r(x) = 1 + x \) and positive integers \( p \). Thus, in the following we will only look at the divisibility properties of binomial coefficients,\(^{19}\) although a similar discussion can be done for general polynomials.\(^{20}\)

In our discussion we will primarily address the question of whether a binominal coefficient is divisible by \( p \) or not. In other words, we consider the black and white coloring of the Pascal triangle interpreted modulo \( p \) (see figure 8.16). The question of divisibility can be approached with the aid of prime number factorization. Below we will see that in order to understand the patterns in Pascal’s triangle formed by the coefficients divisible by an integer \( p \), we should build on the patterns generated by the prime factors of \( p \).

**Pattern in Pascal’s Triangle**

**Mod 3**

Coefficients in the Pascal triangle which are divisible by 3 are shown in black.

We have seen that we can answer the question of divisibility by recursively computing binominal coefficients using an addition rule like eqn. (8.2) modulo \( p \) with a subsequent test as to whether the result is 0 or not. On the other hand, we know very well how to describe the coefficients without a recursion, namely, by

\[
\binom{n}{k} = \frac{n!}{(n-k)!k!}.
\]


Coordinate Systems

Two coordinate systems for the presentation of binomial coefficients. The new modified system is on the right.

The major question for now will be to understand whether or not these coefficients are divisible by $p$ also by means of a direct, nonrecursive computation. It turns out that this problem was solved in a most elegant manner some 150 years ago by the German mathematician Ernst Eduard Kummer. The careful development of Kummer's criterion, which is local in nature, will build a solid foundation for the next step towards understanding the global pattern formation in Pascal's triangle.

It turns out that for the following it will be more convenient to work in a new $(n, k)$-coordinate system (see figure 8.17). The connection between the old and new representation is easy. In the old system we find at position $(n, k)$ the coefficient $\binom{n}{k}$, while in the new system we have at position $(n, k)$ the binomial coefficient

$$\binom{n+k}{k} = \frac{(n+k)!}{n!k!}$$

Figure 8.18 shows the array in the new system, however, rotated and right angled, together with the usual coloration corresponding to even and odd entries.

We will now describe our problem more formally. We define the following set:

$$P(r) = \left\{ (n, k) \mid \binom{n+k}{k} \text{ is not divisible by } r \right\},$$

where $r$ is some integer. Thus, figure 8.18 is a graphical representation of a part of $P(2)$.

Observe that if $p$ and $q$ are two different prime numbers and a given integer is not divisible by $p \cdot q$, then it is also not divisible by $p$ or $q$ alone. Thus,

$$P(pq) = P(p) \cup P(q), \quad \text{if } p \neq q, \ p, q \text{ prime.}$$

For example, $P(6) = P(2) \cup P(3)$ (see figure 8.19). This observation can be generalized to the factorization in prime powers. If we consider the prime factorization of an integer $r$,

$$r = p_1^{s_1} \cdots p_s^{s_s},$$

Divisibility Sets $P(r)$

21E. E. Kummer, Über Ergänzungssätze zu den allgemeinen Reziprozitätsgesetzen, Journal für die reine und angewandte Mathematik 44 (1852) 93–146. For the result relevant to our discussion see pages 115–116.
where the prime numbers $p_k$ are different and the exponents $\tau_k$ are natural numbers, then

$$P(\tau) = P(p_1^{\tau_1}) \cup \cdots \cup P(p_s^{\tau_s}).$$

Thus, to understand the pattern formation for $P(\tau)$ it suffices to understand $P(p^\tau)$,

$$P(p^\tau) = \left\{(n,k) \mid \binom{n+k}{k} \text{ is not divisible by } p^\tau \right\},$$

for a prime number $p$ and some positive integer $\tau$.

Now let us discuss some gems from elementary number theory attributable to Adrien Marie Legendre (1808), Ernst Eduard Kummer (1852), and Edouard Lucas (1877). These results, together with hierarchical iterated function systems, will completely decipher the patterns in $P(p^\tau)$ for any prime number $p$ and positive integer $\tau$.

We would like to have a direct method to check whether $\binom{n+k}{k}$ is divisible by $p^\tau$. Kummer observed that this information is encoded in the $p$-adic representation of $n$ and $k$. You are familiar with decimal expansions like

$$n = d_0 + d_1 \cdot 10 + d_2 \cdot 10^2 + \cdots + d_m \cdot 10^m,$$

where the numbers $d_i \in \{0, \ldots, 9\}$ are the decimal digits. Thus $n$ can be represented as the decimal number

$$n = d_md_{m-1}\ldots d_1d_0.$$
Figure 8.19: The first 66 rows of Pascal’s triangle and its mod-6 pattern generated by a cellular automaton using the rule: \( a_{n,k} = a_{n-1,k} + a_{n,k-1} \mod 6 \).

Now the \( p \)-adic expansion of an integer \( n \) is given analogously by

\[
    n = a_0 + a_1 p + a_2 p^2 + \cdots + a_m p^m
\]

where the digits are now \( a_i \in \{0, \ldots, p - 1\} \) and \( m \) may be different from the \( m \) used above in the decimal number. Corresponding to decimal numbers, we introduce the \( p \)-adic representation

\[
    n = (a_m a_{m-1} \cdots a_1 a_0)_p.
\]
For example, $n = 17$ and $k = 8$ have the expansions

\[
\begin{align*}
n &= (17)_{10} = (10001)_2 = (122)_3 = (32)_5 = (23)_7 = (16)_{11} \\
k &= (08)_{10} = (01000)_2 = (022)_3 = (13)_5 = (11)_7 = (08)_{11}.
\end{align*}
\]

Kummer’s observation concerns the number of carries which occur when we add $n$ and $k$ in the $p$-adic representation when $p$ is a prime. For example, let us add the triadic representations:

\[
\begin{array}{c}
1 & 2 & 2 \\
+0 & 1 & 2 & 2 \\
\hline \\
2 & 2 & 1
\end{array}
\]

Observe that when adding the rightmost digits we obtain a carry to the next digit. The same is true when adding the second digits of the two numbers. Thus, we obtain two carries. On the other hand, if we add the corresponding binary representations:

\[
\begin{array}{c}
10001 \\
+01000 \\
\hline \\
11001
\end{array}
\]

we obtain no carry at all. In other words, if we define

\[c_p(n, k) = \text{number of carries in the } p\text{-adic addition of } n \text{ and } k,\]

we have demonstrated that $c_2(17, 8) = 0$ and $c_3(17, 8) = 2$. Now we can state Kummer’s result.

**Kummer’s Result**

Let $\tau = c_p(n, k)$. \(\binom{n+k}{k}^\tau\) is divisible by the prime power $p^\tau$ but not by $p^{\tau+1}$.

In other words, the prime factorization of \(\binom{n+k}{k}\) contains exactly $c_p(n, k)$ factors of $p$. Applied to our example, $n = 17$ and $k = 8$, we should have that $\binom{n+k}{k}$ has no factors of 2, because $c_2(17, 8) = 0$, and exactly two factors of 3, because $c_3(17, 8) = 2$. Thus, it is an odd number and divisible by 9, but not by 27. In fact, we compute that

\[
\binom{n+k}{k} = \binom{25}{8} = 3^2 \cdot 5^2 \cdot 11 \cdot 19 \cdot 23.
\]

which confirms our conclusions.

Interpreting Kummer’s result the other way around, we conclude from the factorization in eqn. (8.6), that

\[
c_p(17, 8) = \begin{cases} 2, & \text{for } p = 3, 5 \\ 1, & \text{for } p = 11, 19, 23 \\ 0, & \text{otherwise} \end{cases}
\]
In fact, we check, for example, \( c_{11}(17, 8) = 1 \) by adding 17 and 8 in modulo 11 arithmetic, obtaining one carry — as expected,

\[
\begin{array}{c}
16 \\
+ 18 \\
\hline
23.
\end{array}
\]

**Lucas’ Criterion**

To determine whether \( \binom{n}{k} \) is odd or even we can use Lucas’ criterion as follows.\(^{22}\) We compute the binary form of \( n \) and \( k \), say, \( n = 23 = (10111)_2 \) and \( k = 17 = (10001)_2 \). Then we write them one over the other.

\[
\begin{array}{c}
10111 \\
10001 \\
\hline
\end{array}
\]

Now \( \binom{n}{k} \) is odd, if and only if every digit of the bottom number \( k \) is less than or equal to the digit of \( n \) above. This is the case for our example, \( n = 23 \) and \( k = 17 \). In fact, \( \binom{23}{17} = 100047 \) is odd.

Let us see how Lucas’ criterion follows directly from Kummer’s result. Let the binary expansions of \( n \) and \( k \) be

\[
\begin{align*}
 n &= a_m 2^m + a_{m-1} 2^{m-1} + \cdots + a_0 \\
 k &= b_m 2^m + b_{m-1} 2^{m-1} + \cdots + b_0
\end{align*}
\]

with binary digits \( a_i, b_i \in \{0, 1\} \). Since \( k \leq n \) some of the leading binary digits of \( k \) may be 0. First observe that we know from Kummer’s result that \( \binom{n}{k} = \binom{n-k+k}{k} \) is not divisible by 2 if and only if we have for the 2-adic expansion

\[
 n - k = c_m 2^m + c_{m-1} 2^{m-1} + \cdots + c_0
\]

the property

\[
c_i + b_i \leq 1, \quad i = 0, \ldots, m.
\]

In other words, in the \( p \)-adic addition of \( n - k \) and \( k \) there will be no carry.

To complete the argument, we have to show two implications. First, if \( \binom{n}{k} \) is odd, then Lucas’ criterion follows, i.e., \( a_i \geq b_i \), for all \( i \). Second, if Lucas’ criterion is satisfied, then it follows that \( \binom{n}{k} \) is odd.

Let \( i \) denote an arbitrary index \( i \in \{0, \ldots, m\} \). Now we start with the first part, assuming that \( \binom{n}{k} \) is odd. Kummer’s result above states that \( c_i + b_i \leq 1 \). And this implies that \( a_i = b_i + c_i \) and, in conclusion, also \( a_i \geq b_i \), which is what was to be shown. Now we do the second part. We assume \( a_i \geq b_i \). Then \( c_i = a_i - b_i \), which implies that \( c_i + b_i = a_i \leq 1 \). Thus, according to Kummer’s result \( \binom{n}{k} \) is odd. This finishes the second part and completes the proof of Lucas’ criterion.

8.2 Binomial Coefficients and Divisibility

Mod-p Condition

As a particular case we obtain from Kummer’s criterion that \( \binom{n+k}{k} \) is not divisible by the prime number \( p \) provided \( c_p(n, k) = 0 \). In other words, we have

\[
\mathcal{P}(p) = \{(n, k) \mid c_p(n, k) = 0\}.
\]

Moreover, the number of carries \( c_p(n, k) \) is 0 if and only if

\[
a_i + b_i \leq p - 1, \quad i = 0, \ldots, m
\]

where \( a_i \) and \( b_i \) denote the \( p \)-adic digits of \( n \) and \( k \), i.e.,

\[
n = a_0 + a_1 p + a_2 p^2 + \cdots + a_m p^m,
\]

\[
k = b_0 + b_1 p + b_2 p^2 + \cdots + b_m p^m.
\]

This we will call the \textit{mod-p condition}.

For prime powers, Kummer’s result implies that

\[
\mathcal{P}(p^\tau) = \{(n, k) \mid c_p(n, k) < \tau\}
\]

The proof of Kummer’s observation can be based on a formula by Legendre dating from 1808 which determines the largest exponent \( \mu \) of the prime power \( p^\mu \) which divides \( n! \).

---

We recall Kummer’s theorem of 1852.

- Let \( c_p(n, k) \) be the number of carries in the \( p \)-adic addition of \( n \) and \( k \) and \( \tau = c_p(n, k) \). Then \( \binom{n+k}{k} \) is divisible by the prime power \( p^\tau \) but not by \( p^{\tau+1} \).

In order to derive this beautiful result we will use a formula by Legendre from 1808 which deals with the divisibility of \( n! \) by a prime power. The formula is as follows.

- Let \( \mu(n) \) be the largest integer exponent of the prime power \( p^{\mu(n)} \) which divides \( n! = 1 \cdot 2 \cdots n \). Thus, \( n! \) is divisible by \( p^{\mu(n)} \) but not by \( p^{\mu(n)+1} \). Then

\[
\mu(n) = \frac{n - \sigma}{p - 1}
\]

where \( \sigma \) is the sum of the \( p \)-adic coefficients \( a_i \in \{0, 1, \ldots, p - 1\} \) of \( n \),

\[
n = a_0 + a_1 p + a_2 p^2 + \cdots + a_m p^m,
\]

\[
\sigma = a_0 + a_1 + \cdots + a_m.
\]

To show Legendre’s formula we first establish the useful identity

\[
\mu(n) = \sum_{i=1}^{\infty} \left\lfloor \frac{n}{p^i} \right\rfloor,
\]

\[\text{(8.9)}\]
where the brackets \([ \cdot ]\) denote the greatest integer less than or equal to the enclosed quantity. Thus, e.g., \([6.1] = 6\), \([5.9] = 5\), \([\pi] = 3\).

Note that \([n/p^i]\) = 0 for large \(i\). Thus, the sum in eqn. (8.9) is a finite sum. Let us first prove this identity (8.9). We observe that the term \([n/p^i]\) is the number of elements from \(\{1, 2, \ldots, n\}\) which are divisible by \(p^i\). For example, if \(p = 2\), \(i = 3\) and \(n = 17\), then \([n/p^i]\ = 2\). In other words, there are two integers less than or equal to 17 which are divisible by \(2^3\), namely, 8 and 16. Next we observe that the sum in eqn. (8.9) counts any factor in the product

\[1 \cdot 2 \cdot 3 \cdots (n - 1) \cdot n\]

which is divisible by \(p^i\) but not by \(p^{i+1}\) exactly \(i\) times, namely, once in \([n/p]\), once in \([n/p^2]\), \ldots, and once in \([n/p^i]\). This accounts for all occurrences of \(p\) as a factor of \(n!\), i.e., identity (8.9) is established.

Here is an example: \(n = 10\) and \(p = 2\). Thus, we consider factors of 2 in the product \(10! = 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10\). Indeed, 2 is divisible by \(2^1\), 4 by \(2^2\), 6 by \(2^1\), 8 by \(2^3\) and 10 by \(2^1\). Thus, \(10!\) is divisible by \(2^8\). This is shown in the following representation,

\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

\[
p^1 \quad p^2 \quad p^1 \quad p^3 \quad p^1
\]

showing the 8 occurrences of the factor \(p = 2\). Thus, \(\mu(10) = 8\).

Observe that the sum

\[
\sum_{i=1}^{\infty} \left[ \frac{10}{2^i} \right] = \left[ \frac{10}{2} \right] + \left[ \frac{10}{4} \right] + \left[ \frac{10}{8} \right] = 5 + 2 + 1 = 8
\]

is just the number of these occurrences.

Here is another example: \(n = 1000\) and \(p = 3\). 1000! is divisible by \(3^{498}\) but not by \(3^{499}\), since

\[
\mu(1000) = \left[ \frac{1000}{3} \right] + \left[ \frac{1000}{9} \right] + \left[ \frac{1000}{27} \right] + \left[ \frac{1000}{81} \right] + \left[ \frac{1000}{243} \right] + \left[ \frac{1000}{729} \right] = 333 + 111 + 37 + 12 + 4 + 1 = 498.
\]

Let us now establish Legendre's identity. By means of eqn. (8.9) Legendre's formula (8.7) is equivalent to

\[
\sum_{i=1}^{\infty} \left[ \frac{n}{p^i} \right] (p - 1) = n - \sigma. \tag{8.10}
\]

We proceed by showing this relation. Using the \(p\)-adic representation of \(n\) in (8.8) and the definition of the brackets \([ \cdot ]\) we get

\[
\left[ \frac{n}{p^i} \right] = a_i + a_{i+1}p + \cdots + a_mp^{m-i}, \quad i \leq m.
\]
With that we compute the two sums

\[ \sum_{i=1}^{\infty} \left| \frac{n}{p^i} \right| p = \sum_{i=1}^{m} \left( a_i p + a_{i+1} p^2 + \cdots + a_m p^{m-i+1} \right) \]
\[ = a_1 p + a_2 p^2 + a_3 p^3 + \cdots + a_m p^m \]
\[ + a_2 p + a_3 p^2 + \cdots + a_m p^{m-1} \]
\[ + a_3 + \cdots + a_m p^{m-2} \]
\[ \vdots \]
\[ + a_m p \]

\[ \sum_{i=1}^{\infty} \left| \frac{n}{p^i} \right| = \sum_{i=1}^{m} \left( a_i + a_{i+1} p + \cdots + a_m p^{m-i} \right) \]
\[ = a_1 + a_2 p + a_3 p^2 + \cdots + a_m p^{m-1} \]
\[ + a_2 + a_3 p + \cdots + a_m p^{m-2} \]
\[ + a_3 + \cdots + a_m p^{m-3} \]
\[ \vdots \]
\[ + a_m. \]

The difference between the two sums is to be computed,

\[ \sum_{i=1}^{\infty} \left| \frac{n}{p^i} \right| (p - 1) = (a_1 p + a_2 p^2 + \cdots + a_m p^m) \]
\[ - (a_1 + a_2 + \cdots + a_m) \]
\[ = (n - a_0) - (\sigma - a_0) \]
\[ = n - \sigma. \]

This establishes Legendre's identity (8.7).

Now we derive Kummer's criterion from Legendre's identity. Thus, let the \( p \)-adic expansions of \( n \) and \( k \) be

\[ n = a_0 + a_1 p + a_2 p^2 + \cdots + a_m p^m, \]
\[ k = b_0 + b_1 p + b_2 p^2 + \cdots + b_m p^m. \]

where \( a_i, b_i \in \{0, 1, \ldots, p-1\} \). Now, if \( p^\nu \) is the largest prime power of \( p \) which divides \( \binom{n+k}{k} \) then

\[ \nu = \mu(n+k) - \mu(n) - \mu(k), \]

since

\[ \binom{n+k}{k} = \frac{(n+k)!}{n!k!}. \]

In other words, we have to show that

\[ c_p(n, k) = \mu(n+k) - \mu(n) - \mu(k), \quad (8.11) \]
where \( c_p(n, k) \) is the number of carries in the \( p \)-adic addition of

\[
\begin{align*}
    n &= (a_m a_{m-1} \ldots a_1 a_0)_p \quad \text{and} \\
    k &= (b_m b_{m-1} \ldots b_1 b_0)_p.
\end{align*}
\]

Carrying out the addition of these two numbers in the \( p \)-adic representation produces carries \( \epsilon_0, \epsilon_1, \epsilon_2, \ldots \) which are either 0 or 1. Formally, they are obtained from

\[
\epsilon_0 = \left\lfloor \frac{a_0 + b_0}{p} \right\rfloor \quad \text{and} \quad \epsilon_i = \left\lfloor \frac{a_i + b_i + \epsilon_{i-1}}{p} \right\rfloor, \quad i = 1, 2, \ldots
\]

The sum of carries in Kummer’s theorem is

\[
c_p(n, k) = \sum_{i=0}^{\infty} \epsilon_i.
\]

Now we consider the sum \( n + k \) in \( p \)-adic representation, i.e.,

\[
n + k = \sum_{i=0}^{\infty} c_i p^i
\]

where \( c_i \in \{0, 1, \ldots, p-1\} \). If we define for convenience of notation \( \epsilon_{-1} = 0 \) then we can express the \( p \)-adic digits \( c_i \) of \( n + k \) in terms of those of \( n \) and \( k \) and the carries \( \epsilon_i \) as follows:

\[
c_i = a_i + b_i + \epsilon_{i-1} - \epsilon_i p \quad \text{for} \quad i = 0, 1, 2, \ldots
\]

Finally, we use Legendre’s identity to show (8.11):

\[
\nu = \mu(n + k) - \mu(n) - \mu(k) = \frac{n + k - \sum_{i=0}^{\infty} c_i}{p - 1} - \frac{n - \sum_{i=0}^{\infty} a_i}{p - 1} - \frac{k - \sum_{i=0}^{\infty} b_i}{p - 1}
\]

\[
= \frac{1}{p - 1} \left( \sum_{i=0}^{\infty} a_i + \sum_{i=0}^{\infty} b_i - \sum_{i=0}^{\infty} (a_i + b_i + \epsilon_{i-1} - \epsilon_i p) \right)
\]

\[
= \frac{1}{p - 1} \sum_{i=0}^{\infty} \epsilon_i p - \epsilon_{i-1}
\]

\[
= \frac{1}{p - 1} \left( \sum_{i=0}^{\infty} \epsilon_i p - \sum_{i=0}^{\infty} \epsilon_{i-1} \right)
\]

\[
= \frac{1}{p - 1} \left( \sum_{i=0}^{\infty} \epsilon_i p - \sum_{i=0}^{\infty} \epsilon_i \right)
\]

finishing with

\[
\nu = \frac{1}{p - 1} \sum_{i=0}^{\infty} \epsilon_i (p - 1) = \sum_{i=0}^{\infty} \epsilon_i = c_p(n, k).
\]

With this computation, Kummer’s result is established.
8.2 Binomial Coefficients and Divisibility

Coloring Pascal’s Triangle

So far we have only discussed whether a binomial coefficient is divisible by a prime number $p$ or not, which we used for black and white coloring. However, if a coefficient is not divisible by $p$ we could color the respective entry in Pascal’s triangle with one of $p - 1$ colors (depending on the modulus). Let us close this section with a short remark on the computation of the color. The crucial result which determines the color without having to run the associated linear cellular automaton is attributable to Lucas (1877). Let

$$n + k = a_0 + a_1p + a_2p^2 + \cdots + a_mp^m,$$
$$k = b_0 + b_1p + b_2p^2 + \cdots + b_mp^m,$$

where $a_i, b_i \in \{0, \ldots, p - 1\}$ are the $p$-adic digits. Then

$$\binom{n + k}{k} \equiv \binom{a_0}{b_0} \cdot \binom{a_1}{b_1} \cdots \binom{a_h}{b_h} \pmod{p} \quad (8.12)$$

Let us look at an example. According to the above criterion

$$\binom{7}{4} \equiv \binom{1}{1} \cdot \binom{2}{1} \pmod{3}$$

because $7 = (21)_3$ and $4 = (11)_3$. In fact,

$$\binom{7}{4} = 35 \equiv 2 \pmod{3}$$

and also

$$\binom{1}{1} \cdot \binom{2}{1} = 1 \cdot 2 \equiv 2 \pmod{3}.$$

Again, the criterion follows from Legendre’s identity, as did Kummer’s result. We will, however, skip these details and turn now to the description of the global pattern formation in Pascal’s triangle.

In summary, we use the $\text{mod}-p$ condition to test a binomial coefficient for the divisibility by a prime number, and we resort to Lucas’ factorization eqn. (8.12), when we want to know in addition what the value of the coefficient is in the modulo $p$ sense.

---

23 For a proof not using Legendre’s identity, see also N. J. Fine: Binomial coefficients modulo a prime number, Amer. Math. Monthly 54 (1947) 589. Lucas’ identity can also be used to analyze the global structure of the colored Pascal triangle, i.e., the color patterns which are obtained if one uses $p$ colors, one for each modulus ($\equiv 0 \pmod{p}$, $\equiv 1 \pmod{p}$, $\ldots$, $\equiv p - 1 \pmod{p}$). In fact Sved derived Lucas’ result from the geometrical patterns of Pascal’s triangle mod $p$. In other words, the fractal patterns in Pascal’s triangle are equivalent to number theoretical properties of binomial coefficients, and understanding more about the fractal properties will lead to a wider understanding of these number theoretical properties.
8.3 IFS: From Local Divisibility to Global Geometry

You will recall the surprisingly short program ‘Skewed Sierpinski gasket’ from chapter 2. Its secret was hidden in just one BASIC statement:

\[ \text{IF } (x \text{ AND } y) = 0 \text{ THEN PSET } (x+30, y+30) \]

The logical expression of this statement determined whether a point was drawn or not. The expression ‘\(x \text{ AND } y\)’ in the if-clause stands for the bitwise logical AND operation. For example, \(101 \text{ AND } 010\) is 0 while \(101 \text{ AND } 110\) is 1. In other words, the expression is equal to 0 (false) only if no two matching binary digits of \(x\) and \(y\) are both 1. Thus, this expression allows us to test for the occurrence of a carry in the binary addition of the coordinates \(x\) and \(y\). In other words this program uses the Kummer criterion for \(p = 2\) and \(c_2(n,k) = 0\) setting \(n = x\) and \(k = y\). Figure 8.20 gives an impression of the resulting pattern showing more and more details of the Sierpinski gasket. Let us now try to explain why this criterion really is able to generate the Sierpinski gasket.

**Mod-2 Pattern and Binary Addresses**

Address testing on a \(2 \times 2\), \(4 \times 4\) and \(8 \times 8\) grid. More and more details of the familiar Sierpinski gasket are revealed.

Let us consider the unit square in the plane,

\[ Q = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\} \]

Now we expand \(x\) and \(y\) in base 2, i.e.,

\[
\begin{align*}
    x &= \sum_{i=1}^{\infty} a_i 2^{-i}, \quad a_i \in \{0, 1\}, \\
    y &= \sum_{i=1}^{\infty} b_i 2^{-i}, \quad b_i \in \{0, 1\}.
\end{align*}
\]

With this notation we can provide a number theoretical description of the Sierpinski gasket:

\[
S = \{(x, y) \in Q \mid \text{there are binary expansions of } x \text{ and } y \text{ with } a_i + b_i \leq 1, i = 1, 2, \ldots \}
\]

(8.13)
Let us look at some examples:

<table>
<thead>
<tr>
<th>$(x, y)$</th>
<th>$(0,0)$</th>
<th>$(1,0)$</th>
<th>$(1/2,1/2)$</th>
<th>$(3/4,3/4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>0.000...</td>
<td>0.111...</td>
<td>0.1000...</td>
<td>0.11000...</td>
</tr>
<tr>
<td>$y$</td>
<td>0.000...</td>
<td>0.000...</td>
<td>0.0111...</td>
<td>0.10111...</td>
</tr>
<tr>
<td></td>
<td>$\in S$</td>
<td>$\in S$</td>
<td>$\in S$</td>
<td>$\notin S$</td>
</tr>
</tbody>
</table>

The first three points, $(0,0)$, $(1,0)$, and $(1/2,1/2)$ are in $S$. In the last example we also see why we have to say ‘there is an expansion...’ in the characterization of $S$. Otherwise we would not have that $(1/2, 1/2)$ is in $S$. On the other hand $(3/4, 3/4)$ is not in $S$, no matter how we expand, because $a_1$ and $b_1$ must both be 1, i.e., $a_1 + b_1 = 2$.

Note that there is a direct relation to Kummer’s carry-condition above. Indeed, saying that there is a base 2 expansion for $x$ and $y$ with $a_i + b_i \leq 1$, is the same as saying that in adding $x$ and $y$ in the binary number system there is no carry.

In section 5.4 we used iterated function systems (IFS) to convince ourselves that (8.13) indeed characterizes the Sierpinski gasket. We introduced four contractions $w_{00}$, $w_{01}$, $w_{10}$, and $w_{11}$, which contract the unit square $Q$ as in figure 8.21 by a factor of 2:

\[
\begin{align*}
    w_{00}(x, y) &= (x/2, y/2) \\
    w_{01}(x, y) &= (x/2, y/2 + 1/2) \\
    w_{10}(x, y) &= (x/2 + 1/2, y/2) \\
    w_{11}(x, y) &= (x/2 + 1/2, y/2 + 1/2).
\end{align*}
\]

**Figure 8.21**

*Four Similarity Transformations*

Transformations of the square $Q$. 
Three of these transformations \( w_{00}, w_{01}, w_{10} \) provide a Hutchinson equation for the Sierpinski gasket \( S \):

\[
S = w_{00}(S) \cup w_{01}(S) \cup w_{10}(S).
\]  

(8.14)

### The Number Theoretical Description

We base the proof of the number theoretical description in eqn. (8.13) on the definition that the Sierpinski gasket is given by the contractions \( w_{00}, w_{01}, w_{10} \) and the corresponding Hutchinson equation (8.14). Any object (compact, nonempty set) which satisfies this equation must be the Sierpinski gasket because there is only one solution. Thus, to verify that \( S \) from (8.13) is the Sierpinski gasket, we must prove that \( S \) as in (8.13) satisfies (8.14). We proceed by showing the two relations

\[
w_{00}(S) \cup w_{01}(S) \cup w_{10}(S) \subseteq S
\]

and

\[
w_{00}(S) \cup w_{01}(S) \cup w_{10}(S) \supseteq S.
\]

For the first, take any point \((x, y) \in S\) and its binary expansion

\[
(x, y) = (0.a_1a_2 \ldots, 0.b_1b_2 \ldots)
\]

Following eqn. (8.13), \( a_i + b_i \leq 1 \) holds for all indices \( i = 1, 2, \ldots \). Now we apply the three transformations, \( w_{00}, w_{01}, \) and \( w_{10} \),

\[
w_{00}(0.a_1a_2 \ldots, 0.b_1b_2 \ldots) = (0.0a_1a_2 \ldots, 0.0b_1b_2 \ldots)
\]

\[
w_{01}(0.a_1a_2 \ldots, 0.b_1b_2 \ldots) = (0.0a_1a_2 \ldots, 0.1b_1b_2 \ldots)
\]

\[
w_{10}(0.a_1a_2 \ldots, 0.b_1b_2 \ldots) = (0.1a_1a_2 \ldots, 0.0b_1b_2 \ldots)
\]

Clearly, all three resulting points are also in \( S \), because the first digits of the \( x \)- and \( y \)-components of the results are never both equal to 1, and for the remaining pairs of digits the same holds because \( a_i + b_i \leq 1 \) for \( i = 1, 2, \ldots \).

To show the second relation, we again take any point \((x, y) \in S\) with binary expansion as above and have to provide another point in \((x', y') \in S\) such that one of the images \( w_{00}(x', y'), w_{01}(x', y'), \) or \( w_{10}(x', y') \) is equal to the given point \((x, y)\). We may choose

\[
(x', y') = (0.a_2a_3 \ldots, 0.b_2b_3 \ldots)
\]

Note that \( a_1 \) and \( b_1 \) cannot both be equal to 1. Therefore, we immediately obtain

\[
(x, y) = \begin{cases} 
  w_{00}(x', y') & \text{if } a_1 = 0 \text{ and } b_1 = 0 \text{ or } \\
  w_{01}(x', y') & \text{if } a_1 = 0 \text{ and } b_1 = 1 \text{ or } \\
  w_{10}(x', y') & \text{if } a_1 = 1 \text{ and } b_1 = 0
\end{cases}
\]

and there are no other cases. This concludes our proof, and, thus, (8.13) characterizes the Sierpinski gasket.
The binary representation also allows us to see how the iteration of the Hutchinson operator, applied to an arbitrary point in the square $Q$, yields a sequence of points that get closer and closer to the Sierpinski gasket. Observe that if $(x, y) = (0.a_1a_2\ldots, 0.b_1b_2\ldots)$, with arbitrary $a_i$ and $b_i$, then applying the maps $w_{00}, w_{01},$ and $w_{10}$ again and again as in an IFS, yields points with coordinates for which more and more of the leading binary decimals satisfy $a_i + b_i \leq 1$. In other words, starting with

$$A_0 = Q$$

and then running the IFS, generates the sequence

$$A_n = w_{00}(A_{n-1}) \cup w_{01}(A_{n-1}) \cup w_{10}(A_{n-1}), \quad n = 1, 2, \ldots,$$

where the coordinates of the points of $A_n$ satisfy $a_i + b_i \leq 1$ in the leading $n$ binary decimals. Furthermore, the sequence will lead towards the Sierpinski gasket as an attractor, i.e.,

$$A_\infty = S.$$

The first steps are shown in figure 8.22. Now observe that this would be exactly the result of figure 8.20 if the coordinates used in that figure had been preceded by a decimal point. In this case the patterns found on the $2 \times 2$, $4 \times 4$ or $8 \times 8$ grid would exactly match the steps $A_n$ of our iterated function system. But introducing a decimal point in figure 8.20 simply means that we look at rescaled versions of Pascal’s triangle (i.e., scaled by $1/2$, $1/4$ or $1/2^n$ in general). In other words, the mod-2 pattern which we see in Pascal’s triangle is exactly the pattern which we obtain when iterating the IFS which encodes the Sierpinski gasket.

**Divisibility by Primes**

Now we are prepared to look at the patterns obtained from the divisibility of binomial coefficients primes. Or more formally, we want to describe the global pattern formations in

$$P(p) = \left\{ (n,k) \mid \binom{n+k}{k} \text{ not divisible by } p \right\},$$

First we construct an appropriate iterated function system. We consider the unit square $Q$ and subdivide it into $p^2$ congruent squares $Q_{a,b}$ with $a, b \in \{0, \ldots, p - 1\}$. Then we introduce corresponding contraction mappings

$$w_{a,b}(x, y) = \left( \frac{x + a}{p}, \frac{y + b}{p} \right),$$

where

$$w_{a,b}(Q) = Q_{a,b}, \quad a, b \in \{0, \ldots, p - 1\}.$$

This is the generalization of what we have already done for the case $p = 2$ in figure 8.21. Now we define a set of admissible transformations by imposing the restriction

$$a + b \leq p - 1.$$
Patterns of $A_n$

The first three steps of the iterated function system coding the Sierpinski gasket.

![Figure 8.22](image)

This yields a total number of $N = p(p+1)/2$ contractions, each with contraction factor $1/p$. We now introduce the Hutchinson operator $W_p$ corresponding to these $N$ contractions,

$$W_p(A) = \bigcup_{a+b \leq p-1} w_{a,b}(A),$$

where $A$ is any subset of the plane. With the initial set $A_0 = Q$ we can start the iteration

$$A_m = W_p(A_{m-1}), \; m = 1, 2, \ldots$$

and figure 8.23 shows the first two steps for the choice $p = 3$.

In order to keep track of the iteration, we subsequently subdivide each of the $p^2$ subsquares of $Q$ into $p^2$ even smaller ones, and so on repeatedly. Having indexed the first subdivision of $Q$ by $Q_{a,b}$, we continue to label the subsquares of the second subdivision by $Q_{ac,bd}$ and so on. For the example $p = 3$ shown in figure 8.23, the square $Q_{10,12}$ are identified in the following way: the pair $(1,1)$, made from the leading digits in the index of $Q_{10,12}$, determines the center square in the first subdivision, and the pair $(0,2)$ determines the upper left corner square therein. In other words, the square

$$Q_{a_{m-1}a_0b_{m-1}b_0}$$

is a square of the $m^{th}$ generation. We find it by reading the double $p$-adic addresses given by the pair $(a_{m-1}a_0b_{m-1}b_0)$. This natural address-
8.3 From Local Divisibility to Global Geometry

Mod-3 Machine

First two steps of the iteration of the function system $W_p$ for $p = 3$.

Rescaling the Pascal Triangle

Let us now relate the subsquares $Q_{a_m, b_0}$ to the entries of the Pascal triangle. First we generate a geometric model of the divisibility pattern in the Pascal triangle. To this end we equip the first quadrant of the plane with a square lattice so that each square has side length 1. Thus, each square is indexed by an integer pair $(n, k)$ and we call it $R_{n,k}$.

$$R_{n,k} = \{(x, y) \mid n \leq x \leq n + 1, k \leq y \leq k + 1\}.$$ 

The geometrical model of $P(p)$ will now be obtained by selecting all squares $R_{n,k}$ for which $p$ does not divide $\binom{n+k}{k}$:

$$P(p) = \left\{ R_{n,k} \left| \binom{n+k}{k} \text{ is not divisible by } p \right. \right\}.$$ 

We will now relate this infinite pattern to the evolution of the Hutchinson operator, i.e., to the sequence of patterns $A_m$. Note that all $A_m$ are within $Q$ and that $A_m$ is a union of a finite number of squares of side length $1/p^m$. To see the relation between $A_m$ and $P(p)$ we will look at $P(p)$ through a sequence of square ‘windows’ $[0, p^m] \times [0, p^m]$ of side length $p^m$. Now for
The $\mathcal{P}_m(3)$ Subsquares

The squares $\mathcal{P}_1(3)$ (the six black squares with grey underlay in the lower left-hand group) and $\mathcal{P}_2(3)$ (all black squares). Compare with figure 8.23.

Figure 8.24

$m = 1, 2, \ldots$ we pick that part from the geometrical model $\mathcal{P}(p)$ which falls in the corresponding window:

$$\mathcal{P}_m(p) = \mathcal{P}(p) \cap [0, p^m] \times [0, p^m].$$

Figure 8.24 displays $\mathcal{P}_1(p)$ and $\mathcal{P}_2(p)$ for $p = 3$. Comparing $\mathcal{P}_1(p)$ and $\mathcal{P}_2(p)$ with the pattern of $A_1$ and $A_2$ in figure 8.23 we observe that they are identical, though $A_1$ and $A_2$ are in the unit square and $\mathcal{P}_1(p)$ (resp. $\mathcal{P}_2(p)$) fit into a square of side length $p$ (resp. $p^2$). In other words, if we rescale the patterns $\mathcal{P}_m(p)$ by a factor of $1/p^m$ we obtain an object which we want to show is identical with $A_m$. To this end we introduce

$$\mathcal{S}_m(p) = \frac{1}{p^m} \cdot \mathcal{P}_m(p)$$

or more explicitly

$$\mathcal{S}_m(p) = \left\{ \frac{z}{p^m} \mid z \in \mathcal{P}_m(p) \right\}.$$

Indeed, each subsquare in $\mathcal{S}_m(p)$ is indexed by an integer pair $(n, k)$ such that $p$ does not divide $\binom{n+k}{k}$. In other words each such subsquare is identical with $Q_{a_{m-1} \ldots a_0, b_{m-1} \ldots b_0}$, where $n = (a_{m-1} \ldots a_0)_p$ and $k = (b_{m-1} \ldots b_0)_p$ according to Kummer’s mod-$p$ condition. Summarizing we have that

$$A_m = \mathcal{S}_m(p), \ m = 1, 2, \ldots$$
8.3 From Local Divisibility to Global Geometry

The First Primes

The limit sets of the rescaled geometric models of \( P(2) \), \( P(3) \), \( P(5) \), and their associated IFSs (i.e., a graphical representation of the transformations \( \omega_{ab} \)).

Figure 8.25

As we let \( m \) go to infinity, we know that \( A_m \) will converge towards the attractor of the IFS, and consequently the rescaled geometric models \( S_m(p) \) will also converge to the attractor of the IFS.\(^{24}\) We denote the limit set by \( S(p) \). In this manner we have just seen that the rescaled geometric models have a limit set. It represents a rescaled geometric model of the (infinite) Pascal triangle modulo \( p \) which we denoted by \( P(p) \).\(^ {25}\)

Figure 8.25 shows the resulting geometric models \( S(p) \) when running the IFSs corresponding to \( P(p) \) for \( p = 2, 3, 5 \). The approach by iterated function systems allows us to compute the fractal dimensions of these sets:

<table>
<thead>
<tr>
<th></th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S(2) )</td>
<td>( \log 3 / \log 2 \approx 1.585 )</td>
</tr>
<tr>
<td>( S(3) )</td>
<td>( \log 6 / \log 3 \approx 1.631 )</td>
</tr>
<tr>
<td>( S(5) )</td>
<td>( \log 15 / \log 5 \approx 1.683 )</td>
</tr>
<tr>
<td>( S(7) )</td>
<td>( \log 28 / \log 7 \approx 1.712 )</td>
</tr>
</tbody>
</table>

If \( p \) is prime then the formula for the self-similarity dimension of \( S(p) \) is

\[
D_s = \frac{\log p(p + 1)/2}{\log p}.
\]

Can you explain why? Just note, that these sets are strictly self-similar and recall the definition of the self-similarity dimension from chapter 4.

\(^{24}\)Convergence is with respect to the Hausdorff metric.

\(^{25}\)In this regard we also refer to S. J. Willson, *Cellular automata can generate fractals*, Discrete Applied Math. 8 (1984) 91–99 who studied limit sets of linear cellular automata via rescaling techniques.
8.4 HIFS and Divisibility by Prime Powers

We have seen that iterated function systems are in some sense the natural framework in which to decipher the global pattern formation obtained by the divisibility properties in Pascal’s triangle (or of pattern formations in linear cellular automata). We have, however, only taken the first step, namely, with respect to the divisibility by a prime $p$. Our next step, considering divisibility by prime powers $p^r$, is rather long compared with the first one. We will describe how in this case the global patterns arising in Pascal’s triangle can be completely understood through hierarchical IFSs. The fractal patterns in

$$P(p^r) = \left\{ (n, k) \mid \binom{n+k}{k} \text{ is not divisible by } p^r \right\}$$

can be deciphered by a hierarchical IFS whose design and properties are in very close correspondence with Kummer’s criterion for finding the largest prime power which divides $\binom{n+k}{k}$. Figure 8.26 shows two examples where we observe that the straightforward self-similarity properties of the sets $P(p)$ have been replaced by hierarchies of self-similarity features.

**First Prime Powers**

Rescaled geometric models of $P(4)$ and $P(8)$.

Figure 8.26

Let us start with some observations concerning the example $P(4)$. In this case Kummer’s criterion implies:

$$P(4) = \{ (n, k) \mid c_2(n, k) = 0 \text{ or } c_2(n, k) = 1 \}.$$  

If $c_2(n, k) = 0$, then $\binom{n+k}{k}$ is not divisible by 2. If $c_2(n, k) = 1$, then $\binom{n+k}{k}$ is divisible by 2, but not by 4. Thus, if either one of the conditions is satisfied, then $\binom{n+k}{k}$ is not divisible by 4. Therefore, we also have to take into account those coordinates $(n, k)$ whose $p$-adic addition have exactly one carry (i.e., where $\binom{n+k}{k}$ is divisible by 2 but not by 4). How can we reflect this property in an iterated function system?

---


27 See section 5.9.
8.4 HIFS and Divisibility by Prime Powers

The graph of a hierarchical iterated function system for the mod-4 example. Iterating this system generates three images, one for each node (the nodes representing three networked MRCMs). The final image of node 1 is a Sierpinski gasket. The final image of node 3 is the desired mod-4 pattern.

Again we look at the unit square $Q$ and the contractions $w_{00}, \ldots, w_{11}$ of figure 8.21.

We have seen that $m$ iterations of these transformations, first applying $w_{a_0b_0}$ then $w_{a_1b_1}$, etc. lead to the subsquare

$$w_{a_{m-1}b_{m-1}}(\cdots w_{a_0b_0}(Q)) = Q_{a_{m-1}\cdots a_0,b_{m-1}\cdots b_0}.$$

So far we have only considered the case where the binary addition of $(a_{m-1} \cdots a_0)_2$ and $(b_{m-1} \cdots b_0)_2$ had no carry (i.e., $a_i + b_i \leq 1$). Now, how can we achieve exactly one carry? At first glance it appears that all we have to use is the transformation $w_{11}$. For example, applying the sequence $w_{01}, w_{10}, w_{11}, w_{01}, w_{00}$ would lead to

$$Q_{0011001101} = w_{00}(w_{01}(w_{11}(w_{10}(w_{01}(Q))))).$$

And indeed in this example the addition would provide a carry in the third binary decimal. But wait; there is also a carry in the fourth binary decimal counted from the right! What is wrong? Obviously we have to be a bit more careful. The transformation $w_{11}$ provides a carry — so far okay — but the transformation which follows has to be $w_{00}$. Otherwise we would obtain another carry. Therefore, having $w_{11}$, followed by $w_{01}$, as in our example, is not allowed.

Figure 8.27 shows the graph of a hierarchical iterated function system which reflects our observations. The nodes 1, 2 and 3 represent three networked MRCMs. The first one operates in a feedback loop applying the Hutchinson operator $w_{00} \cup w_{01} \cup w_{10}$. The second one transforms the output of the first one using $w_{11}$. The third machine again operates in a feedback
loop with \( w_{00} \cup w_{01} \cup w_{10} \), but additionally it merges the output of the second machine, transformed by \( w_{00} \).

How do we iterate this network? We start the iteration with three copies of the unit square \( Q \), one for each node, \( A_0(1) = Q \), \( A_0(2) = Q \), and \( A_0(3) = Q \). The first step provides for these nodes:

\[
A_1(1) = w_{00}(Q) \cup w_{01}(Q) \cup w_{10}(Q) = Q_{0,0} \cup Q_{0,1} \cup Q_{1,0} \\
A_1(2) = w_{11}(Q) = Q_{1,1} \\
A_1(3) = w_{00}(Q) \cup w_{00}(Q) \cup w_{01}(Q) \cup w_{10}(Q) = Q_{0,0} \cup Q_{0,1} \cup Q_{1,0}
\]

and for the second step we obtain

\[
A_2(1) = \bigcup_{a+b \leq 1} w_{a,b}(Q_{0,0} \cup Q_{0,1} \cup Q_{1,0}) \\
= Q_{00,00} \cup Q_{00,01} \cup Q_{01,00} \\
\cup Q_{00,10} \cup Q_{00,11} \cup Q_{01,10} \\
\cup Q_{10,00} \cup Q_{10,01} \cup Q_{11,00} \\
A_2(2) = w_{11}(Q_{0,0} \cup Q_{0,1} \cup Q_{1,0}) = Q_{10,10} \cup Q_{10,11} \cup Q_{11,10} \\
A_2(3) = \bigcup_{a+b \leq 1} w_{a,b}(Q_{0,0} \cup Q_{0,1} \cup Q_{1,0}) \cup w_{0,0}(Q_{1,1}) \\
= Q_{00,00} \cup Q_{00,01} \cup Q_{01,00} \\
\cup Q_{00,10} \cup Q_{00,11} \cup Q_{01,10} \\
\cup Q_{10,00} \cup Q_{10,01} \cup Q_{11,00} \cup Q_{01,01}
\]

These steps and the next are visualized in figure 8.28. In step \( m \) we obtain in node 1 all subsquares \( Q_{a_{m-1} \ldots a_0,b_{m-1} \ldots b_0} \) whose indices produce no carry, which yields the rescaled geometric model \( S_m(2) \). For node 2 we obtain subsquares with the carry produced by the leading binary decimal of its indices, and for node 3 we obtain subsquares whose indices provide no carry or just one. In other words, \( A_m(3) \) provides the desired geometric model of \( P(4) \) after rescaling by \( 1/2^m \) (see also figure 8.30).

Let us now build a hierarchical iterated function system for general prime powers \( p^\tau \). As before, we consider contractions with contraction factor \( 1/p \)

\[
w_{a,b} : Q \to Q, \text{ where } w_{a,b}(Q) = Q_{a,b} 
\]

and \( w_{a,b}(x, y) = ((x + a)/p, (y + b)/p) \) for \( a, b \in \{0, \ldots, p-1\} \). Again the unit square \( Q \) is subdivided into \( p^2 \) congruent squares \( Q_{a,b} \), which are indexed by the \( p \)-adic pair \((a, b)\).

Our hierarchical IFS will have \( 2\tau - 1 \) nodes (i.e., there will be \( 2\tau - 1 \) individual IFSs which are networked with each other). Each IFS will use mappings from definition 8.15, which operate on a unit square \( Q \). In order to distinguish them, we designate them \( Q^1, \ldots, Q^{2\tau-1} \).
Figure 8.28: The first three steps of the networked MRCM for $P(4)$ shown in figure 8.27.

In principle, a hierarchical IFS can have mappings from any $Q^i$ to any other $Q^j$. However, our particular hierarchical IFS will only have particular connections. They are the systematical extension of what we already have seen for $P(4)$. Figure 8.29 shows the resulting network. The black squares are numerated from 1 to $2^r - 1$ and represent the individual nodes of our HIFS. The arrows between squares specify Hutchinson mappings. More precisely, we have four types of such mappings. They are distinguished by an index to the letter $W$, which determines a selection of the contraction mappings $w_{a,b}$. More precisely, let $B$ be any (compact) subset of the plane and $a,b \in \{0, \ldots, p-1\}$. Then:

$$
W_{\leq p-1}(B) = \bigcup_{a+b \leq p-1} w_{a,b}(B), \\
W_{\leq p-2}(B) = \bigcup_{a+b \leq p-2} w_{a,b}(B), \\
W_{\geq p-1}(B) = \bigcup_{a+b \geq p-1} w_{a,b}(B), \\
W_p(B) = \bigcup_{a+b \geq p} w_{a,b}(B).
$$

In our hierarchical IFS we have

$$
W_{\leq p-1} : Q^{2i-1} \rightarrow Q^{2i-1}, \quad i = 1, \ldots, \tau, \\
W_{\geq p} : Q^{2i-1} \rightarrow Q^{2i}, \quad i = 1, \ldots, \tau - 1, \\
W_{\geq p-1} : Q^{2i} \rightarrow Q^{2i+2}, \quad i = 1, \ldots, \tau - 2, \\
W_{\leq p-2} : Q^{2i} \rightarrow Q^{2i+1}, \quad i = 1, \ldots, \tau - 1.
$$

Now we run this IFS. Starting with a unit square for each node, we see an evolution of patterns in each node. In fact, as we run the machine sufficiently long, the evolution of these patterns will begin to stagnate, i.e., run into a
Tower Machine

Tower of IFSs and their network channels.

Figure 8.29

limit. Imagine that what we have termed the tower machine in figure 8.29 has viewing windows through which we can watch this evolution. Our particular interest would be to monitor the evolution in the nodes shown on the right-hand side (i.e., the nodes corresponding to $Q^{2k-1}, k = 1, \ldots, \tau$). The result is that we see exactly the patterns which are given by $P(p^k)$: the pattern in node $2k - 1$ corresponds to the pattern in $P(p^k)$.

In particular the global pattern of $P(p^\tau)$ can be found in $Q^{2\tau - 1}$. But also note that the network in figure 8.29 shows how we have to mix together all the $P(p^k), 1 \leq k \leq \tau - 1$ to obtain $P(p^\tau)$. In other words, the hierarchical IFS not only allows the generation of the pattern of $P(p^\tau)$, even more importantly, it exactly deciphers the hierarchy of the self-similarity features in $P(p^\tau)$.

Our result once again shows very strongly that hierarchical IFSs are not just there to make pretty pictures. They are deeply rooted in pure mathematics. They appear here as entirely natural for the explanation of the discussed geometrical patterns in the Pascal triangle.
Let us spend a little more effort to explain how the tower machine in figure 8.29 actually runs. We construct a \((2\tau - 1) \times (2\tau - 1)\) matrix \(F_{p^\tau}\). The entries in this matrix are two kinds of symbols, the empty symbol \(\emptyset\), or one of the Hutchinson mappings from 8.16. Let us denote the general element of \(F_{p^\tau}\) by \(f_{ij}\), i.e., we think of this element as a mapping from \(Q^j\) to \(Q^i\). Thus, if according to figure 8.29 there is no mapping, we put the empty symbol:

\[
F_{p^\tau} = (f_{ij}), \quad i, j = 1, \ldots, 2\tau - 1.
\]

Then

\[
\begin{align*}
    f_u &= W_{\leq p-1}, \text{ whenever } l \text{ is odd}, \\
    f_{l+1, l} &= W_{\geq p}, \text{ whenever } l \text{ is odd}, \\
    f_{l+1, l} &= W_{\leq p-2}, \text{ whenever } l \text{ is even}, \\
    f_{l+2, l} &= W_{\geq p}, \text{ whenever } l \text{ is even}, \\
    f_{l+1, l} &= W_{\geq p}, \text{ whenever } l \text{ is odd}, \\
    f_{kl} &= \emptyset, \quad \text{ otherwise.}
\end{align*}
\]  

(8.17)

Let us look at the example of \(\tau = 3\). Here we have a \(5 \times 5\) matrix:

\[
\begin{pmatrix}
    W_{\leq p-1} & \emptyset & \emptyset & \emptyset & \emptyset \\
    W_{\leq p} & \emptyset & \emptyset & \emptyset & \emptyset \\
    \emptyset & W_{\geq p-2} & W_{\leq p-1} & \emptyset & \emptyset \\
    \emptyset & W_{\geq p-1} & W_{\geq p} & \emptyset & \emptyset \\
    \emptyset & \emptyset & \emptyset & W_{\leq p-2} & W_{\leq p-1}
\end{pmatrix}
\]

Now we let this matrix \(F_{p^\tau}\) operate on a stack of subsets of the plane, say \(B = (B_1, \ldots, B_{2\tau-1})\), and let \(C = F_{p^\tau}(B)\), i.e., if \(C = (C_1, \ldots, C_{2\tau-1})\), then

\[
C_k = \bigcup_{l=1, \ldots, 2\tau-1} f_{kl}(B_l)
\]

where \(f_{kl}(B_l) = \emptyset\), provided \(f_{kl} = \emptyset\).

Now let \(A_0 = (Q^1, \ldots, Q^{2\tau-1})\) and \(A_{m+1} = F_{p^\tau}(A_m), m = 0, 1, 2, \ldots\). We can analyze the content of component \(2s - 1, s = 1, \ldots, \tau\), in the stack of objects \(A_m\) for each \(m\) and compare it with a rescaled geometrical model of a colored part of the Pascal triangle, which we call \(S_m(p^s)\), as in the analysis of divisibility by \(p\). More precisely we introduce unit squares at a lattice point \((n, k)\) in \(\mathbb{R}^2\)

\[
R_{n,k} = \{(x, y) \in \mathbb{R}^2 | x = n + u, y = k + v, (u, v) \in Q\}
\]

representing a black cell in the colored version of \(\mathcal{P}_m(p^s)\), i.e., an entry \((n, k)\) in \(\mathcal{P}_m(p^s)\) for which \(\binom{n+k}{k}\) is not divisible by \(p^s\). If we rescale \(R_{n,k}\) by \(1/p^m\) we obtain a square (in the unit square \(Q\)) which has width \(1/p^m\), i.e., a square which can be identified with one of the squares \(Q_{a_{m-1} \ldots a_0, b_{m-1} \ldots b_0}\), where \(n = (a_{m-1} \ldots a_0)_p\) and \(k = \ldots \).
Figure 8.30: The top node of this IFS network generates the Sierpinski gasket $S_2$. The upper three nodes form the part for $S_4$ (compare with figures 8.27 and 8.28). The whole network generates $S_8$.

We will write alternatively $Q_{n,k}$. Then we introduce the rescaled geometrical model for the colored version of $P_m(p^s)$:

$$S_m(p^s) = \bigcup_{n,k} Q_{n,k}, \text{ where } 0 \leq n + k \leq p^m - 1 \text{ and } \binom{n+k}{k} \text{ not divisible by } p^s.$$ 

The collection of little squares in $S_m(p^s)$ will be identical with the $(2s-1)^{th}$ component of $A_m$. Figure 8.30 shows the results in the hierarchical IFS for $p = 2$ and $\tau = 3$. The layout is adopted from that in figure 8.29.

To complete the argument we use Kummer’s criterion, according to which $\binom{n+k}{k}$ is not divisible by $p^s$ if $s > c_p(n,k)$, where $c_p(n,k)$ is
Plate 15: 3-dimensional cross section of a Julia set in 4-dimensional quaternion space, © R. Lichtenberger.

Plate 16: Different view of the same Julia set with cut open 2-dimensional cross section revealing the corresponding Julia set in the complex plane, © R. Lichtenberger.
Plate 17: Julia set of the quadratic family for \( c = -11 + 0.67i \). This is close to a parabolic situation.

Plate 18: Julia set of the quadratic family \( x^2 + c \). For \( c = -0.39054 - 0.58679i \) a Siegel disk is obtained.

Plate 19: 3D-rendering of the potential of a connected Julia set.
Plate 20: Four different renderings of one detail of the Mandelbrot set. The coloring of the first image (top) is computed by the escape-time-method and corresponds to equipotential lines. The 3D-rendering (middle) shows the potential of the Mandelbrot set. This image is the cover of the book *The Beauty of Fractals*. The distance-estimator rendering (bottom left) uses colors to represent the distance to the Mandelbrot set while the 3D-rendering (bottom right) shows height corresponding to distance.

Plate 21: High resolution image of the potential of a piece of the Mandelbrot.
Plate 22: Natural ice formation on Mount Kilimanjaro, © John Reader.

Plate 23: Two trajectories on the Lorenz attractor with color indicating distance to unstable steady states.

Plate 25: Original enlargement of the Mandelbrot set used for the rendering in plate 24, cover of *Scientific American*, August 1985.

Plate 26: Variation of the rendering in plate 24.
Plate 27: The pendulum experiment from section 12.8. The basins of attraction of the three magnets are colored red, blue, and yellow.

Plate 28: Detail of plate 27 showing the intertwined structure of the three basins.
the number of carries in the \( p \)-adic addition of \( n \) and \( k \). Observe that each arrow in figure 8.29 is marked by a 0 or 1 which represents the carry.

Now we can complete the comparison of the \((2s - 1)\)th component of \( A_m \) in the iteration of the hierarchical IFS and the rescaled part of the Pascal triangle \( S_m(p^s) \). Any of the little squares in \( A_m \) is of the form \( Q_{a_{m-1} \ldots a_0, b_{m-1} \ldots b_0} \), where the pair \((a_{m-1} \ldots a_0, b_{m-1} \ldots b_0)\) satisfies \( c_p(a_{m-1} \ldots a_0, b_{m-1} \ldots b_0) \leq s - 1 \) by the construction of the hierarchical IFS (see figure 8.29). Indeed, the mappings of type \( W_{\leq p-1} \) and \( W_{\leq p-2} \) will produce no carry, while a mapping of type \( W_{\geq p} \) will obviously produce a carry. But also a mapping of type \( W_{\geq p-1} \) in figure 8.29 will produce a carry because it will always be preceded by a mapping of type \( W_{\geq p} \). Finally, we note that

\[
\begin{align*}
n &= a_0 + a_1 p + \cdots + a_{m-1} p^{m-1}, \\
k &= b_0 + b_1 p + \cdots + b_{m-1} p^{m-1}.
\end{align*}
\]

This shows that the two patterns in the \((2s - 1)\)th component of \( A_m \) and the rescaled model of \( S_m(p^s) \) agree. As a consequence of the above observations, we also obtain that the rescaled geometric models have a limit set \( S(p^s) \) as \( m \to \infty \).

Let us finally remark that there is an illuminating formula for the fractal dimension of these objects appearing in the components of a limit set of a hierarchical IFS.\(^{28}\) Applying this formula to our example we obtain that the Hausdorff dimension of \( S(p^\tau) \) is equal to the Hausdorff dimension of \( S(p) \), i.e., it is independent of \( \tau \). This result has been obtained in a different way by John M. Holte by exploiting Kummer’s result.\(^{29}\) Intuitively the independence of the Hausdorff dimension of \( S(p^\tau) \) from \( \tau \) is suggested from the images of \( S(p^\tau) \), where we can observe that the patterns in \( P(p^\tau) \) are in some hierarchical fashion just mixtures of the patterns for \( P(p) \).


8.5 Catalytic Converters, or How Many Cells Are Black?

Pascal's triangle has been in existence for many centuries and has inspired beautiful investigations. We have seen the first step of how it has laid the foundation for the understanding of pattern formation for linear cellular automata in one dimension.\textsuperscript{30} But it has also recently sparked the investigation of a problem which at first glance seems to have no relation to the triangle at all.

Assume we were to play darts with a large Pascal triangle as a target. What are the probabilities that we would hit a black cell or a white one, or more precisely, an odd or an even number, or a number which is divisible by 3 or one which is not, or a number which is divisible by $p^r$ or one which is not? Our discussion of the global patterns in the Pascal triangle makes it possible to answer such questions. We just have to evaluate the corresponding areas in the structures corresponding to the rescaled geometric models of $P(p^r)$. Depending on the parameters, this may turn out to be a rather technical computation.

Let us look at a related question which allows a more immediate answer. We again use the original coordinate system for the Pascal triangle (see the left option of figure 8.17). How many black cells are there in the $r^{th}$ row? In other words, how many of the numbers which appear in the $r^{th}$ row are not divisible by 2, or 3, or 5, or any other prime number $p$? There is a remarkably direct procedure to arrive at the result. First, take $r$ and expand it with respect to base $p$,

$$r = c_0 + c_1 p + c_2 p^2 + \cdots + c_m p^m, \quad c_i \in \{0, \ldots, p-1\}.$$ 

Now let $h_p(r)$ be defined by

$$h_p(r) = \prod_{i=0}^{m} (c_i + 1) = (c_0 + 1) \cdot (c_1 + 1) \cdots (c_m + 1). \quad (8.18)$$

Then $h_p(r)$ is the number of entries in the $r^{th}$ row of the Pascal triangle which are not divisible by $p$.

**Determining the Count $h_p$**

Let us give an argument using again the modified coordinate system as in figure 8.17 (right). Here the $r^{th}$ row is characterized by $n + k = r$. Thus, we are asking for the cardinality of

$$\left\{(n, k) \mid n + k = r \text{ and } \binom{n + k}{k} \text{ is not divisible by } p \right\}.$$ 

Consider the $p$-adic representations of $n$, $k$, and $r$,

$$n = a_m p^m + \cdots + a_1 p + a_0, \quad a_i \in \{0, \ldots, p-1\},$$

$$k = b_m p^m + \cdots + b_1 p + b_0, \quad b_i \in \{0, \ldots, p-1\},$$

$$r = c_m p^m + \cdots + c_1 p + c_0, \quad c_i \in \{0, \ldots, p-1\}.$$ 

What is the probability of hitting a point of the mod-2 pattern in the Pascal triangle?

According to Kummer's mod-$p$ criterion, \( \binom{n+k}{k} \) is not divisible by $p$ if and only if $a_i + b_i \leq p - 1$ for all $i = 0, \ldots, m$. In this case there is no carry in the $p$-adic addition of $n$ and $k$. Thus, the coefficients of the sum $r = n + k$ must satisfy

$$c_i = a_i + b_i, \text{ where } i = 0, \ldots, m.$$  

How many choices of $n$ are there such that this condition is satisfied? For the $i$th $p$-adic digit there are $c_i + 1$ such choices, namely, $a_1 = 0, \ldots, c_i$. Thus, the total number of possible choices is the product

$$h_p(n) = \prod_{i=0}^{m} (c_i + 1).$$

Exactly that many entries $(n, k)$ in row $r$ have the property that $\binom{n+k}{k}$ is not divisible by $p$.

Connection to the Invariant Measure

Figure 8.32 shows $h_p(r)$ for $p = 2$ as a function of $r$. You might recognize this graph. It seems that we already obtained the same function in our discussion of the invariant measure for the chaos game (see figure 6.25). Let us give a short argument for this coincidence. Observe that for $p = 2$ the count of black cells in row $r + 2^m$ (with $r < 2^m$) is $h_p(r + 2^m) = 2h_p(r)$ (compare equation 8.18, where $(c_m + 1) = 2$). Thus, it is twice as large as the count in row $r$. In other words, if we look at the first $2^{m+1}$ rows of Pascal’s triangle with mod-2 coloring, 1/3 of the black cells fall into the first $2^m$ rows and 2/3 fall into the second $2^m$ rows, and this is true for all $m = 1, 2, \ldots$ This observation links the count $h_p(r)$ to the construction of the invariant measure in figure 6.25: the density plot $h_k$ corresponds exactly to the count $h_p(r)$, for $0 < r < 2^k$. A detailed discussion of this measure (and the various connections to the chaos game and Pascal’s triangle) are carried out in the appendix on multifractal measures.
Number of Entries
The number of entries in the $r\text{th}$ row of the Pascal triangle.

![Figure 8.32](image)

Reaction Rate Measurement
Chemical reaction rate in a catalytic oxidation process.

![Figure 8.33](image)

Now imagine the graph of figure 8.32 flipped over and compare with figure 8.33, which shows the measurement of the chemical reaction rate as a function of time in a catalytic oxidation process. The remarkable resemblance of the flipped-over graph and this kind of measurements provided the motivation to model a catalytic converter by one-dimensional cellular automata.\(^\text{31}\) Thus, we are back to the relation of cellular automata and polynomials which we discussed at the beginning of this chapter. This relation allows us to interpret $h_p(r)$ as a count of “oxidized” cells in an appropriate cellular automaton. In this sense our discussion has provided a first glimpse of an idea of why modeling a catalytic converter by cellular automata could be a successful approach and exhibits the qualitative behavior found in real chemical experiments.