# 18.821 PROJECT 3: FINITE STRINGS FOUND IN ROWS OF PASCAL'S TRIANGLE MOD M

ROGERS EPSTEIN, BEN MYERS, AND CONNOR SELL

## 1. Introduction

Pascal's triangle, known for its numerous convenient combinatorial properties, has a well-studied structure. However, when we consider the elements of Pascal's triangle taken mod $m$ for some integer $m$, some surprising structures arise. For instance, taken mod 2, Pascal's triangle looks very much like a Sierpinski triangle (see Figure 6). Analogous self-similarity occurring when $m$ is prime also gives rise to some interesting results.

In this paper we address the problem of determining which strings of integers appear consecutively in some row of the triangle modulo different $m$'s. We give some definitions to better study this question:

**Definition 1.** *Let $S = (a_1, \ldots, a_n)$ be a length $n$ string of integers mod $m$. Then we say that Pascal's triangle mod $m$* contains *$S$ if $S$ occurs consecutively in some row.*

**Definition 2.** *Let $P_m$ denote the multiset of (finite-length) strings contained in Pascal's Triangle mod $m$. And $P_m^k$ will denote the multiset of strings contained in Pascal's Triangle mod $m$ up to and including row $k$.*

*We are primarily interested in the size of the set of distinct strings $\{s \in P_m \mid |s| = n\}$ over different $m$s and $n$s. We call this number $a_n(m)$.*

We have conjectured certain formulas for cases where we fix $n = 2$ or $m = 2$. First, we conjecture $a_2(m) = m^2$. We've made some progress towards proving these conjectures, including proving that $a_2(p) = p^2$ for prime $p$. Additionally, we show that it is possible to find a string consisting of arbitrarily many zeros in Pascal's Triangle mod $m$ for any $m$.

Second, we conjecture $a_n(2) = n^2 - n + 2$. This has proven more difficult, but we are able to give some insight into the structure of Pascal's triangle mod 2 which could be useful for further research.

We also have a result which is helpful in using computers to compute $a_n(p)$ for $p$ a prime.

**Theorem** (Search Bound Theorem)**.** *Suppose $p$ is prime, and let $p^d$ be the smallest power of $p$ such that $p^d \geq n - 1$. Then all strings of length $n$ that occur in Pascal's triangle mod $p$ appear in the first $p^{d+2}$ rows.*

Last, after we defined the density of strings in Pascal's Triangle mod a prime $p$, we found the following theorem.

**Theorem** (Density of Zero Strings)**.** *For all lengths $n$ and moduli $m$, the string of $n$ zeros mod $m$ has density* 1*.*

## 2. Heuristic Results

The following gives the values of $a_n(m)$. The blank spaces were too computationally intensive.

| n\m | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 9 | 16 | 25 | 36 | 49 | 64 | 81 |
| 3 | 8 | 25 | 56 | 101 | 200 | 253 | 393 | 513 |
| 4 | 14 | 43 | 116 | 169 | 598 | 403 | 848 | 943 |
| 5 | 22 | 71 | 190 | 253 | 1515 | 589 | 1420 | 1729 |
| 6 | 32 | 109 | 283 | 353 | | 811 | 2115 | 2653 |
| 7 | 44 | 157 | 401 | 483 | | 1069 | 2981 | 3751 |
| 8 | 58 | 207 | 532 | 643 | | 1363 | 3950 | 4865 |

In order to study $a_n(m)$, we wrote a program. Using it, we found that for $n = 2$ and $m \leq 30$, that $a_n(m) = m^2$, which means that all possible strings of length 2 are contained in Pascal's Triangle mod $m$.

**Conjecture 1.** *For all $m$, $a_2(m) = m^2$.*

## 3. Pascal's Triangle Mod a Prime

One statement about $a_n(p)$ that we have been able to prove conclusively is that $a_n(p) = p^2$ when $n = 2$ and $p$ is prime. In other words, every possible string of length 2 appears in Pascal's triangle modulo any prime $p$. To understand why this is the case, we first investigate the structure of Pascal's triangle modulo a prime.

It is enlightening to look at some $m$th row of the triangle, when $m$ is divisible by $p$. We can find all the values for $\binom{m}{k}$ in two distinct cases - when $k$ is divisible by $p$, and when it's not divisible by $p$.

**Proposition 1.** Case 1: *If $k$ is divisible by $p$, then we can write $\binom{m}{k} = \binom{pi}{pj}$ for some $i$ and $j$. And we have*

$$\binom{pi}{pj} \equiv \binom{i}{j} \mod p$$

Case 2: *If $k$ is not divisible by $p$, then $\binom{m}{k} \equiv 0 \mod p$.*

*Proof. Case 1:* We shall prove this combinatorially. Imagine we have a grid on the surface of a cylinder, with $p$ columns and $i$ rows. Note that there is no "leftmost" or "rightmost" column, but there is a "top" and "bottom" row. Suppose we want to analyze the number of unique ways to fill in $pj$ of the $pi$ boxes in the grid. An upper bound for this number is simply $\binom{pi}{pj}$, but this obviously overcounts most cases. We will first try to find how many cases there are for which this method does not double them. This is the case when every row is completely filled or empty. Then, any rotation of the configuration would look the same, so we would have only counted it once. All these cases are determined simply by choosing the $j$ rows of the possible $i$ to fill, meaning there are $\binom{i}{j}$ such configurations.

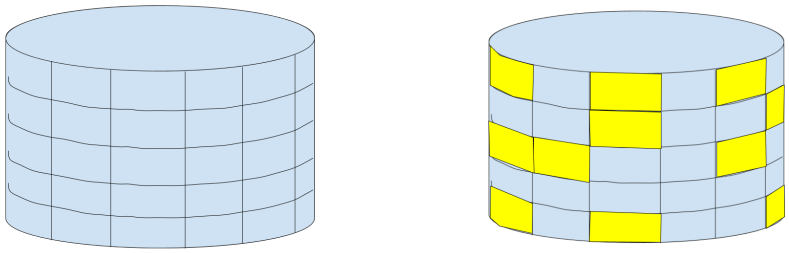

FIGURE 1. On the left, a square grid on the surface of a cylinder. On the right, some squares are filled in.

In all other cases, we can rotate the cylinder $p$ times to get different configurations. This is true because there is at least one row not fixed under the rotation operation, and since there are a prime number of columns, the order of this operation must be $p$. So, we double counted all of these cases by a factor of $p$. This means that

$$\binom{pi}{pj} - \binom{i}{j} \equiv 0 \mod p$$

This is precisely what we aimed to prove.

*Case 2:* This case is even easier. It is impossible for any of the $\binom{m}{k} = \binom{pi}{k}$ configurations to be fixed since it would have to be the case that every row is either filled or empty, which contradicts that $p$ does not divide $k$. Thus, all configurations in this case are counted $p$ times, and so $\binom{m}{k} \equiv 0 \mod p$, as desired.

□

Now we turn our attention to some self-similarities in Pascal's triangle mod $p$. The $pi$th row of Pascal's triangle mod $p$ looks like row $i$ with $p-1$ zeros crammed in between each two entries, as demonstrated in Figure 2. We can easily see that each nonzero entry has $p-1$ zeros on either side, since every nonzero element has index divisible by $p$. Now we can use our intuition for these rows divisible by $p$ to fill out the rest of the triangle.

Recall that each number in Pascal's triangle is the sum of the two numbers above it. Precisely, $\binom{m}{k} = \binom{m-1}{k-1} + \binom{m-1}{k}$. This means that each entry depends entirely on the two entries above it. But note that each of these two entries depend entirely on the two entries above them, so our original entry $\binom{m}{k}$ can be expressed solely in terms of $\binom{m-2}{k-2}$, $\binom{m-2}{k-1}$, and $\binom{m-2}{k}$. In fact, because this dependence is linear, we can show inductively that any arbitrary $\binom{m}{k}$ is a particular linear combination of $\binom{m-i}{k-i}$, $\binom{m-i}{k-i+1}$, $\ldots$, $\binom{m-i}{k}$.

In particular, our nonzero entry $\binom{pi}{pj}$ in the $pi$th row is the only nonzero entry in some linear combination we have for $\binom{pi+x}{pj+y}$, where $0 \leq y < x \leq p-1$. The way Pascal's triangle grows within this region must be identical to the way Pascal's triangle grows

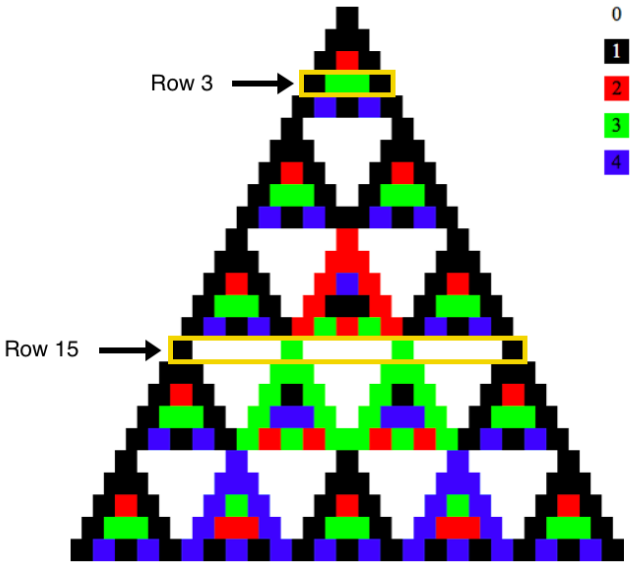


FIGURE 2. Row $15 = 3 \times 5$ of Pascal's triangle mod 5 is similar to Row 3. They contain the same entries, but Row 15 has 5 - 1 = 4 zeros in between each of Row 3's entries.

from $\binom{0}{0}$ for $p-1$ rows, up to some multiple. This is because the relevant entries in the $pi$th row are a multiple of the corresponding entries in the zeroth row, by virtue of having only one nonzero entry. Thus, for the next $p-1$ rows and for each nonzero entry in row $pi$, we have a triangle that's a multiple of the first $p$ rows of Pascal's triangle, multiplied by $\binom{pi}{pj}$. Figure 2 provides a visual for this result.

Also note that when we take row $i$ and put $p-1$ zeros between each of its entries to get row $pi$, some of the entries in row $i$ may already be zeros. In fact, whenever we consider the $m$th row, with $m = p^2 i$ for some integer $i$, then every $p$th entry comes from row $pi$, which already has $p-1$ consecutive zeros. Then any nonzero entries $\binom{p^2 i}{k}$ must have $k$ divisible by $p^2$, and so we have $p^2 - 1$ consecutive zeros. We can inductively show, in fact, that in any $(p^d i)$th row, the only entries $\binom{p^d i}{k}$ that might be nonzero are the ones for which $p^d$ divides $k$.

This means we can get $p^d - 1$ zeros in a row. And since each entry depends entirely on the entries above it, each of these rows of $p^d - 1$ zeros form an upside-down triangle that extends $p^d - 1$ rows down. In particular, there are $\frac{(p^d-1)(p^d)}{2} = \frac{1}{2}(p^{2d} - p^d)$ zeros in every $p^d$ by $p^d$ block – that's $\frac{1}{2} - \frac{1}{2p^d}$ of the entries.

We can also apply similar logic as above, in which we showed that the entries below a nonzero entry with $p-1$ zero entries on either side generates a multiple of the first $p$ rows of Pascal's triangle. This time, since we have $p^d - 1$ zero entries on either side, it'll be longer before any nonzero entries can interfere with the triangle's development, so we get a multiple of the first $p^d$ rows of Pascal's triangle. In particular, the first $p^{d+1}$ rows consist of zeros and triangles that are multiples of the first $p^d$, with the bases starting at each $ip^d$. And the first $p^{d+2}$ rows have multiples of the first $p^{d+1}$ rows for

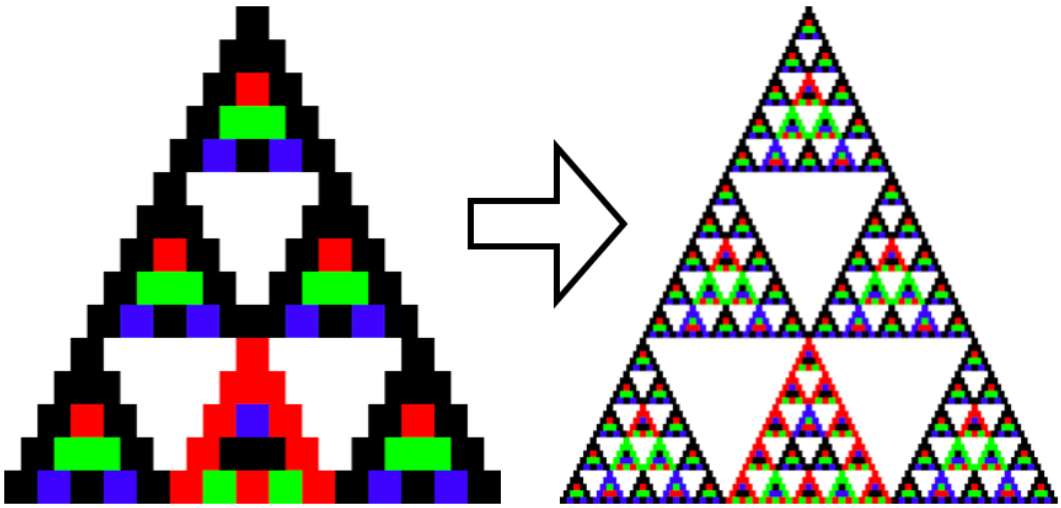

FIGURE 3. The first 15 and then the first 75 rows of Pascal's triangle mod 5. Notice how each entry on the left corresponds to a triangle of side length 5 on the right.

their nonzero elements, and so on. If you look closely at Figure 3, you'll notice that the triangle on the left is the top part of each of the six large triangles on the right, except for the bottom middle one which is doubled.

Now we're ready to prove something about the values of $a_n(m)$ for prime $m$.

**Theorem 1.** *For all primes $p$, $a_2(p) = p^2$. In other words, every valid string of length 2 appears somewhere in Pascal's triangle modulo $p$.*

*Proof.* First, note that the first entry in each row is $\binom{i}{0} = 1$. The second entry is $\binom{i}{1} = i \pmod p$, so the second entry cycles through the $p$ different values.

Next, recall that starting with $\binom{pi}{pj}$, we have the first $p$ rows of Pascal's triangle mod $p$, multiplied by $\binom{pi}{pj} = \binom{i}{j}$, as demonstrated by Figure 3. For any length 2 string $(z_1, z_2)$ such that $z_1$ and $z_2$ are both nonzero, we can consider $\frac{z_2}{z_1}$, because $\mathbb{F}_p$ is a field when $p$ is prime. We know we can find $(1, \frac{z_2}{z_1})$ as the first two entries in row $\frac{z_2}{z_1}$, which must be within the first $p$ rows of Pascal's triangle, and we know that $\binom{pz_1}{p} = \binom{z_1}{1} = z_1$, so $\binom{pz_1}{p}$ serves as the base of a triangle equivalent to the first $p$ rows of Pascal's triangle multiplied by $z_1$. Thus, at position $\binom{pz_1+\frac{z_2}{z_1}}{p}$ and $\binom{pz_1+\frac{z_2}{z_1}}{p+1}$, we can find $z_1(1, \frac{z_2}{z_1}) = (z_1, z_2)$. We can use this argument any time that $z_1$ and $z_2$ are both nonzero, so any pair of nonzero entries can be found in Pascal's triangle. (Note that $z_2$ must be nonzero as well as $z_1$ even though it's not in the denominator since $(1, 0)$ doesn't necessarily appear until row $p$, which is not one of the first $p$ rows.)

The case in which $z_1$ or $z_2$ is zero is simpler. Recall that $\binom{pz}{p} = \binom{z}{1} = z$, but $\binom{pz}{p+1}$ and $\binom{pz}{p-1}$ are both zero since $p+1$ and $p-1$ are not divisible by $p$. Thus at this point we can find both $(0, z)$ and $(z, 0)$. In fact, if $z = p \equiv 0 \pmod p$, we can find $(0, 0)$ as well. Thus any pair of entries containing a zero appears in Pascal's triangle.

We can now find any pair that does or doesn't contain a zero. Thus, every length 2 string mod $m$ appears in Pascal's triangle mod $m$, as desired. □

We don't have a proof for the exact values of $a_n(p)$ for larger $n$ with $p$ prime, but we do have a surprising bound on the values. Since the number of possible strings mod $p$ of length $n$ is $p^n$, one might expect $a_n(p)$ to grow exponentially for fixed $p$. However, for prime $p$, we can prove this is not the case.

**Theorem 2.** *For a fixed prime $p$, the sequence $a_n(p)$ is bounded above by a quadratic in $n$.*

*Proof.* To prove this statement, we will first prove a lemma. The lemma states that all strings of length $n$ that occur in Pascal's triangle mod $p$ will occur in the first several rows, and gives a bound on how many rows need to be checked to find all the appearing strings. The linearity of the bound allows us to constrain the number of strings that can appear.

This lemma is also useful on its own – it was introduced earlier as the Search Bound Theorem.

**Lemma 1** (Search Bound Theorem)**.** *Suppose $p$ is prime, and let $p^d$ be the smallest power of $p$ such that $p^d \geq n-1$. Then all strings of length $n$ that occur in Pascal's triangle mod $p$ appear in the first $p^{d+2}$ rows.*

*Proof.* We'll use proof by induction on $d$. To restate the theorem in terms of $d$, every string of length $n \leq p^d + 1$ that occurs in Pascal's triangle occurs in the first $p^{d+2}$ rows.

For the base case, consider $d = 0$. We need to show that for $n \leq p^0 + 1 = 2$, any string of length $n$ that occurs in Pascal's triangle occurs in the first $p^2$ rows. But in our proof of Theorem 1, we gave explicit locations where each string of length 2 could be found, and they all occurred in the first $p^2$ rows. And any string of length less than 2 occurs as a substring of a string of length 2, and thus must also occur within the first $p^2$ rows. Our lemma holds for $d = 0$.

Now for the induction step. Consider arbitrary $d \geq 1$ , and suppose the induction statement holds for $d-1$. Recall that we can construct the first $p^{d+2}$ rows of Pascal's triangle by replacing each entry in the first $p^{d+1}$ rows of Pascal's triangle with the first $p$ rows, including the zeros to the right, multiplied by the value of the entry – see Figure 3. Thus, we can view our $p^{d+2}$ rows as $p^{d+1}$ distinct "chunks" of size $p$ by $p$ that are entirely determined by the corresponding value in the first $p^{d+1}$ rows. Figure 4 depicts these chunks, and gives an example of how a string can stretch across multiple chunks.

Note that any string of length $n \leq p^d + 1$ can stretch across $p^{d-1} + 1$ chunks at most. Since each chunk in the first $p^{d+2}$ rows is determined by a single value from the first $p^{d+1}$ rows (see Figure 3), each string of length $n$ is determined by a string of length $n' \leq p^{d-1} + 1$. But by our induction assumption, all strings of length $n' \leq p^{d-1} + 1$ that appear in Pascal's triangle appear in the first $p^{d+1}$ rows. Thus every set of $p^{d+1}$ chunks that appears at all appears in the first $p^{d+2}$ rows, so the same applies for strings of length $n \leq p^d + 1$. This completes our induction step. □

Now, notice that the first $p^3 n$ rows will always contain all strings of length $n$ that appear at all. Note that the first inequality holds because if $pn < p^d + 1$, then $n <$

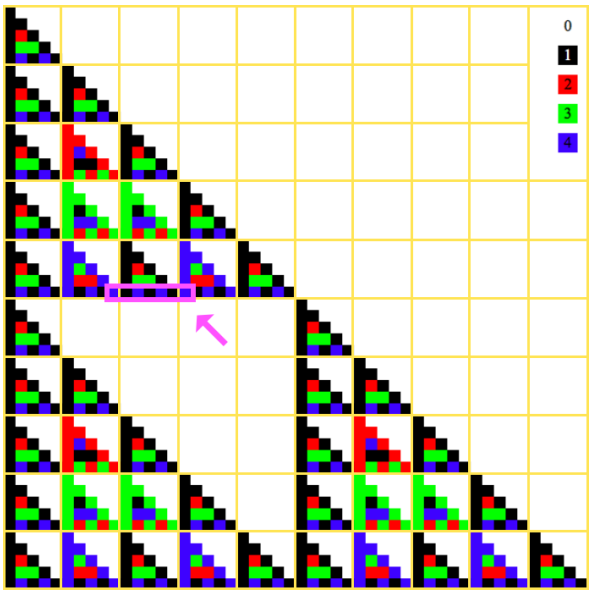


FIGURE 4. Pascal's triangle mod 5 has been left-justified to make clear the locations of the chunks, demarcated by the yellow lines. The pink arrow points to a string of length $7 = 5 + 2$ stretching across 3 different chunks.

$p^{d-1} + \frac{1}{p} < p^{d-1} + 1$, so we chose the wrong $d$.

$$p^3 n = p^2(mn) \geq p^2(p^d + 1) = p^{d+2} + p^2 > p^{d+2}$$

Thus, for fixed $p$, the number of rows we have to check to find all strings is linear in $n$. But there are no more than $p^3 n$ entries in any one row, so there are at most $(p^3 n)^2 = p^6 n^2$ places for a string to start. Thus no more than $p^6 n^2$ different strings can occur in the first $p^3 n$ rows, and since no new strings can occur further on, we have that $a_n(p) \leq p^6 n^2$. In particular, if we fix $p$, we can see that $a_n(p)$ is bounded by a quadratic in $n$. This is exactly what we wanted to show. □

**Corollary 1.** *For a fixed squarefree integer $m$, the sequence $a_n(m)$ is bounded above by a polynomial in $n$. This polynomial has degree $2k$, where $k$ is the number of prime factors of $m$.*

*Proof.* Suppose our squarefree integer $m$ has prime factorization $p_1 p_2 \dots p_k$. By Theorem 2, for each $n$ and each $i$, we have $a_n(p_i) \leq p_i^6 n^2$. For any string of length $n$ occurring in Pascal's triangle mod $m$, the corresponding string must of course appear in Pascal's triangle mod $p_i$ for each $i$. By the Chinese Remainder Theorem, then, the maximum number of strings of length $n$ that occur mod $m$ is given by

$$a_n(m) \leq a_n(p_1) a_n(p_2) \dots a_n(p_k) = (p_1^6 n^2)(p_2^6 n^2) \dots (p_k^6 n^2) = m^6 n^{2k}$$

Thus for fixed, squarefree $m$, $a_n(m)$ is bounded above by a polynomial, and the degree of the polynomial is twice the number of prime factors of $m$, as desired. □

So not only is $a_n(p)$ polynomially bounded for fixed primes $p$, but $a_n(m)$ is polynomially bounded in general for fixed squarefree integers $m$. In fact, we conjecture that $a_n(m)$ is polynomially bounded for all fixed $m$.

## 4. Frequency of Zeros

When studying $P_m$ in order to gain insight into $a_m(n)$, it soon becomes clear that certain strings are easier to find than others. In particular, zeros are especially common. This insight led us to the following simple theorem:

**Theorem 3.** *For any length $n$ and modulus $m$, it is possible to find a string of $n$ zeros in $P_m$.*

*Proof.* Consider the string $\binom{n!\cdot m}{i}, i = 1, 2, \ldots, n$. We can see $\binom{n!\cdot m}{i} = \frac{n!\cdot m}{i!(n!\cdot m-i)!} = \frac{n!\cdot m\cdot\ldots}{i!} \equiv 0$, since $i!|n!$ for all $i \leq n$, and $m|m$. □

Thus, for all $m, n$, we can always find a string mod $m$ of $n$ zeros. While it is interesting to prove whether or not a given string of length $n$ exists in $P_m$, one could also ask how frequently this string appears in $P_m^k$ relative to other strings of length $n$. This motivates us to define a density:

**Definition 3.** *For a given string $s$ of length $n$, let*

$$\sigma_m^n(s) = \lim_{k\to\infty} \frac{\#\textit{times string } s \textit{ appears in } P_m^k}{\#\textit{not necessarily distinct strings of length } n \textit{ in } P_m^k}.$$

We assume that for all strings $s$ and $m, n \in \mathbb{N}$, $\sigma_m^n(s)$ should exist. Under this assumption, we are curious what values $\sigma$ can take, and for what string inputs. We have shown the following theorem, with the proof following:

**Theorem 4.** *For all lengths $n$ and moduli $m$, the string of $n$ zeros mod $m$ has density 1.*

From the previous section, we know that Pascal's Triangle mod a prime $p$ has a fractal-like structure. In particular, the triangle up through the $p^k$th row looks like a bunch of copies of the triangle up through the $p^{k-1}$th row, up to multiplication by a constant, along with upside-down triangles of all zeros. In particular, for any prime $p$, this gives us a lower bound on the intermediate density of 0's up through a row of the form $p^k$. This results from the fact that from this self-similarity we can find a constant $0 < c_p < 1$ such that at least a proportion of $c_p$ of the numbers in Pascal's Triangle mod $p$ are zeros, since the upside-down triangles between the smaller copies of itself are made up of entirely zeros. In particular, in each iteration precisely $\frac{p(p-1)}{2}$ of the $p^2$ triangles are "copies" of the original one, and since $p \geq 2$, this shows that $c_p \geq \frac{p-1}{2p} = \frac{1}{2} - \frac{1}{2p} \geq \frac{1}{4}$.

Then, since the density $d_p$ of $\sigma_p^1(0)$ exists, we can say the intermediate values through each row of the form $p^k$ will converge to this density, and thus must satisfy the property that $d_p \geq (1 - c_p)d_p + c_p \cdot 1$, or $d_p \geq 1$. Since $d_p \leq 1$, we know $\sigma_p^1(0) = 1$ for all prime $p$.

To show this for moduli that are prime powers $p^k$, we use Kummer's Theorem, which is as follows:

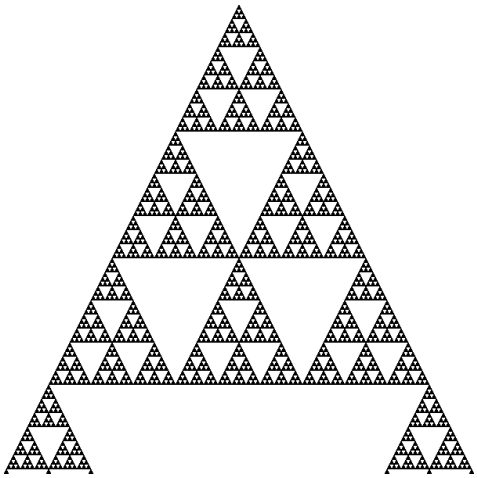

FIGURE 5. For $p = 3$, we can see that the first $3^k$ rows look just like the first $3^{k+1}$. Note that we can check $c_p \geq \frac{1}{2}$.

**Theorem 5.** ***Kummer's Theorem:*** *The largest integer $k$ such that $p^k$ divides the binomial coefficient $\binom{m}{n}$ is equal to the number of carries that occur when $n$ and $m-n$ are added in the base $p$.*

Then, we aim to show that for a given positive integer $k$ and prime $p$, "most" choices of positive integers $n$ and $r = m - n$ create a sum with at least $k$ carries in base $p$. The following is an outline of a proof for this: We can lower bound the number of carries when adding $n$ and $r$ with the number of corresponding digits between them in base $p$ that add to at least $p$. Note that there are $\sum_{i=0}^{p-1} i = \frac{p(p-1)}{2}$ possible pairs of such digits out of the total $p^2$ choices. In particular, for all primes $p$, at least $\frac{1}{4}$ of these pairs will sum to at least $p$. Since this probability is positive and constant, then for arbitrarily longer and longer numbers numbers, the probability of $k$ pairs adding to at least $p$ goes to 1. Thus, as the digits of $r$ and $n$ go to infinity, the probability that there are at least $k$ digits goes to 1, as desired.

It is not hard to extend these results. For all moduli $m$, there cannot be a nonzero element $x \in \mathbb{Z}/m\mathbb{Z}$ such that $\sigma_m^1(x) > 0$, since if there were, we could mod out the triangle by a prime power divisor of $m$, $p^k$, that does not divide $x$ and demonstrate that $\sigma_{p^k}^1(x \mod p^k) > 0$, which is a contradiction. Thus, we've proven our theorem for $n = 1$.

Now, suppose that for a given $n > 1$ and modulus $m$, there is a string $s$ that had density $d > 0$. Then, for each element $x \in s$, $\sigma_m^1(x) > d/n > 0$. However, this is only true for $x = 0$, so the only such string $s$ for which this is possible is that of $n$ zeros. Since this string is the only one with nonzero density, its density must be 1. □

So our theorem is proven. Note the following corollary follows immediately:

**Corollary 2.** *For all lengths $n$ and moduli $m$, any string of length $n$ containing a nonzero element has density 0.*

## 5. Mod 2

As in the general case of Pascal's triangle mod a prime, visual symmetries are immediately apparent in Pascal's triangle mod 2, as is easily seen in Figure 6. However, the strings in $P_2$ have interesting properties, but we have unable to provide an explicit description of them.

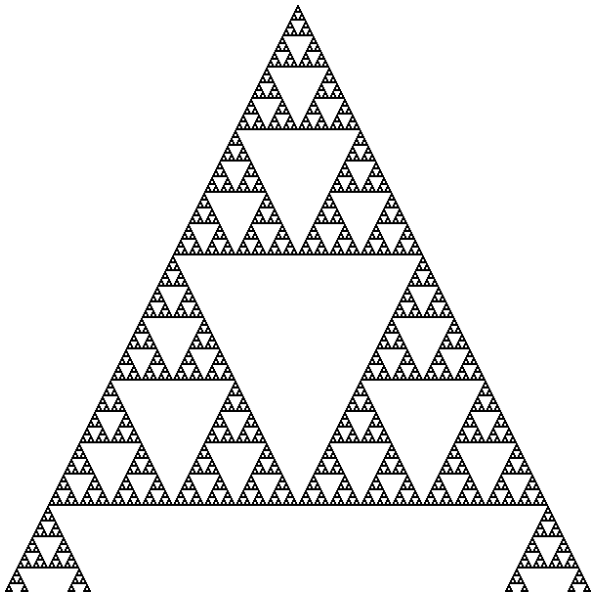

Figure 6. Pascal's Triangle Mod 2 with black ones and white zeros. Notice the resemblance to the Sierpinski triangle.

Consider the following data computed via a Python program (and applying Theorem 1). We include $2^n$, which is the total number of binary strings of length $n$.

| n | $a_n(2)$ | $2^n$ |
|---|---|---|
| 1 | 2 | 2 |
| 2 | 4 | 4 |
| 3 | 8 | 8 |
| 4 | 14 | 16 |
| 5 | 22 | 32 |
| 6 | 32 | 64 |
| 7 | 44 | 128 |
| 8 | 58 | 256 |
| 9 | 74 | 512 |
| 10 | 92 | 1024 |
| 11 | 112 | 2048 |
| 12 | 134 | 4096 |
| 13 | 158 | 8192 |

As is easily checked, this is consistent with the following conjecture.

**Conjecture 2.** *For all $n \in \mathbb{N}$, $a_n(2) = n^2 - n + 2$.*

We've made some progress towards understanding which strings occur in $P_2$. Some patterns are easily discernible. First, we have a result providing some families of strings which are in $P_2$, in addition to what we know from Theorem 3 that all strings consisting solely of zeros are in $P_2$. See Figure 6 for the intuition behind the theorem, noticing the rows consisting of all ones and the rows before and after them.

**Theorem 6.** *The following families of strings are found in $P_2$:*

*(a) Strings consisting only of ones.*
*(b) $(1, 0, 1, 0, 1, \ldots, 0, 1)$: strings of alternating ones and zeros.*
*(c) $(1, 0, 0, \ldots, 0, 1)$: strings consisting of a 1 followed by $2^n$ zeros, for any $n$, then followed by another 1.*

*Proof.* (a). We will prove something slightly stronger, in fact. Namely, a row consists only of ones if and only if it is the $(2^n - 1)$th row for some non-negative integer $n$. The desired result will follow immediately.

By Lucas's Theorem (see Theorem 8 in the Appendix),

$$\binom{m}{k} \equiv \prod_{i=0}^{n} \binom{m_i}{k_i} \mod p$$

where $m_i, k_i$ are the binary digits of $m$ and $k$, respectively. Notice that an immediate corollary is that $\binom{m}{k}$ is divisible by a prime 2 if and only if at least one of the binary digits of $k$ is greater than the corresponding binary digit of $m$.

So, letting $m = 2^n - 1$, we see that $m$ has binary representation $1 \cdots 1$, so each digit is maximal. Thus, corresponding digits of $k$ cannot be larger. By the Corollary, $\binom{m}{k}$ is not divisible by 2.

Conversely, if $m \neq 2^n - 1$, then one of its digits is $m_i = 0$, so we can pick $k = 1 \cdot 2^i$. Then by Lucas's Theorem, $\binom{m}{k}$ is even, so the row has at least one 0.

(b) follows from (a) since such a row necessarily precedes a row of all ones.

(c) also follows from (a) since such a row necessarily follows a row of all ones. □

We have also nailed down some necessary conditions for strings in $P_2$. Before we state these, we make some simple definitions.

**Definition 4.** *We call a maximal-length substring consisting only of ones or only of zeros a* run.

For instance, $(0, 0, 0, 1, 1, 1)$ contains a run $(1, 1, 1)$ while $(1, 1)$ is not a run. Also, $(0, 1, 1, 1, 1, 0)$ contains two runs of the form $(0)$ and a run of the form $(1, 1, 1, 1)$. Clearly, it is possible to decompose any binary string into its runs.

**Definition 5.** *An* interior run *in a string $S$ is a run which doesn't touch either end of $S$.*

For instance, $(0, 0, 0, 1)$ contains no interior runs, but $(1, 0, 0, 0, 1)$ contains $(0, 0, 0)$ as an interior run.

**Definition 6.** *Let $f(n)$ be defined to be the unique power of 2 such that $f(n) \mid n$, and $2f(n) \nmid n$.*

**Theorem 7.** *Let $S$ be a binary string such that $S \in P_2$. Let $L$ be the set of lengths of interior runs in $S$. Then*

*(a) Any interior runs of ones must have length $2^k$ for some $k$.*
*(b) For any lengths $l_1, l_2 \in L$, $f(l_1) = f(l_2)$.*

First, we prove a lemma about two short strings not in $P_2$.

**Lemma 2.** *The strings $(1,1,0,1), (1,0,1,1) \notin P_2$.*

*Proof.* By Theorem 1, it suffices to show that $(1,1,0,1)$ is not found in $P_2^{16}$, the first 16 rows of Pascal's Triangle mod 2. This has been checked using a computer program. □

*Proof of Theorem.* **(a)** We use induction on the rows of Pascal's Triangle. The base case is easy to check. Suppose that up to the $j$th row, we only see strings with interior runs of ones of lengths $2^k$. As you can check, it suffices to consider the following cases where strings of consecutive ones might occur in the next row:

Case 1: Row $j$ contains $S = (\ldots, x, x, 1, 0, 1, 0, \ldots, 1, y, y, \ldots)$.
Case 2: Row $j$ contains $S = (\ldots, 1, 1, 0, 1, \ldots)$.
Case 3: Row $j$ contains $S = (\ldots, 0, 0, 1, 1, 1, \ldots, 1, 0, \ldots, 0)$, where the run is of length $2^k$.

*Case 1:* This is the meaty case of the proof. We will show that such a string does not occur unless the substring consisting of alternating ones and zeros has length $2^n - 1$.

First note that if $x \geq 1$ or $y \geq 1$, we know the string is not in $P_2$ by Lemma 2 since $(1,1,0,1)$ or $(1,0,1,1)$. Thus, we can assume $x = y = 0$. In this case, the structure of the triangle–as shown in Figure 6 and proven in Section 3–will guarantee that the substring $(1,0,1,0,\ldots,1,0,1)$ is a previous row of Pascal's triangle which we've already dealt with inductively or an entire new row. If all of the string $S$ were a substring of a previous row and didn't have length $2^n - 1$, then this would contradict our induction assumption. As shown in the proof of Theorem 6, an entire row consisting of alternating ones and zeros occurs exactly at rows $2^n - 2$ for each $n$, which has length $2^n - 1$, as desired.

*Case 2:* That this cannot occur follows immediately from Lemma 2.

*Case 3:* The row following this, by the principal that an element in Pascal's Triangle is determined by adding the elements above it, will be $(\ldots, 0, 1, 0, \ldots, 0, 1, 0)$, which satisfies the condition still.

This proves part (a).

**(b)** First, note that if we regard the first $2^k$ rows of Pascal's triangle, its structure is as in Figure 7. This self-similarity within Pascal's triangle follows from Proposition 1 but can also be explained using Theorem 6 part (a) for the following reason: after we have a row of all ones, we have a row of the form $(1,0,0,\ldots,0,1)$ at which point the first portion of Pascal's triangle repeats itself. This structure, with three identical triangles and an upside-down triangle in the center, will allow us to perform induction

on the rows, showing that within each row, the runs of ones and zeros with lengths satisfy the given condition. It will suffice to prove this since any string in $P_2$ is the substring of some row (by definition of $P_2$) and thus has interior runs which are also interior runs of some row.

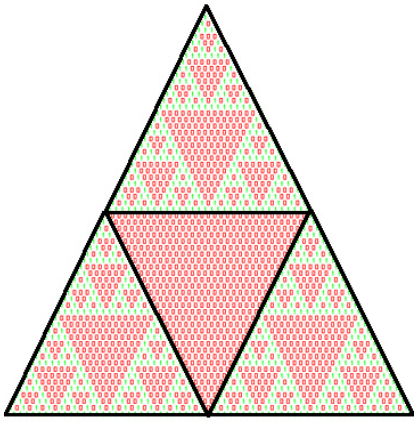

FIGURE 7. Consider the first $j = 63$ rows of Pascal's Triangle. Notice the upside-down triangle of zeros at the center and the self-similarity property.

Before we perform the induction, we need only establish two easy facts about the structure of rows.

(i) Let $2^{k-1} \leq j < 2^k$. Then the run of zeros at the center of row $j > 2$ has length $z_j = 2^{\lceil \log_2(j) \rceil} - j - 1$.
(ii) In row $j$, the length of any run consisting of ones is $x_j = f(j+1)$.

Without going into detail, (i) follows by simply observing where the upside-down triangle of zeros occurs, which starts on row $2^k$ and has length decreasing by one as we go from row $j$ to row $j+1$. Similarly (ii) could be argued via induction on rows, using the proof of Theorem 6 part (a) and the self-similarity properties discussed.

Now, assuming $z_j \neq 0$, we know that for any $m$

$$2^m | x_j \;\Leftrightarrow\; 2^m | j+1 \;\Leftrightarrow\; 2^m | -j-1 \;\Leftrightarrow\; 2^m | (2^l - j - 1),$$

where $2^m | 2^l$. Let $2^l = 2^{\lceil \log_2 j \rceil}$. Then since exactly those powers of two dividing $x_j$ divide $z_j$, we conclude that $f(x_j) = f(z_j)$.

Now we're ready to use (strong) induction, aiming to show that for any lengths $l_1, l_2$ of runs in a given row, $f(l_1) = f(l_2)$. The base case is clear, looking at the rows up to, say, row 3. Now suppose the hypothesis is true for all rows $k$ for $k < j$. We will show that it is also true for row $j$.

Using the structure of Pascal's triangle discussed, we can decompose the row $j$ into three strings $S_1, S_2, S_3$, where $S_2$ is the run of zeros at the center of the row. $S_1$ is identical to $S_3$, and both are identical to row $k = j - 2^{\lfloor j \rfloor}$ of Pascal's triangle mod 2. This comes from the fact that we have three identical triangles and one triangle of zeros, as highlighted in Figure 8. Thus, within $S_1$ and $S_3$, the induction hypothesis

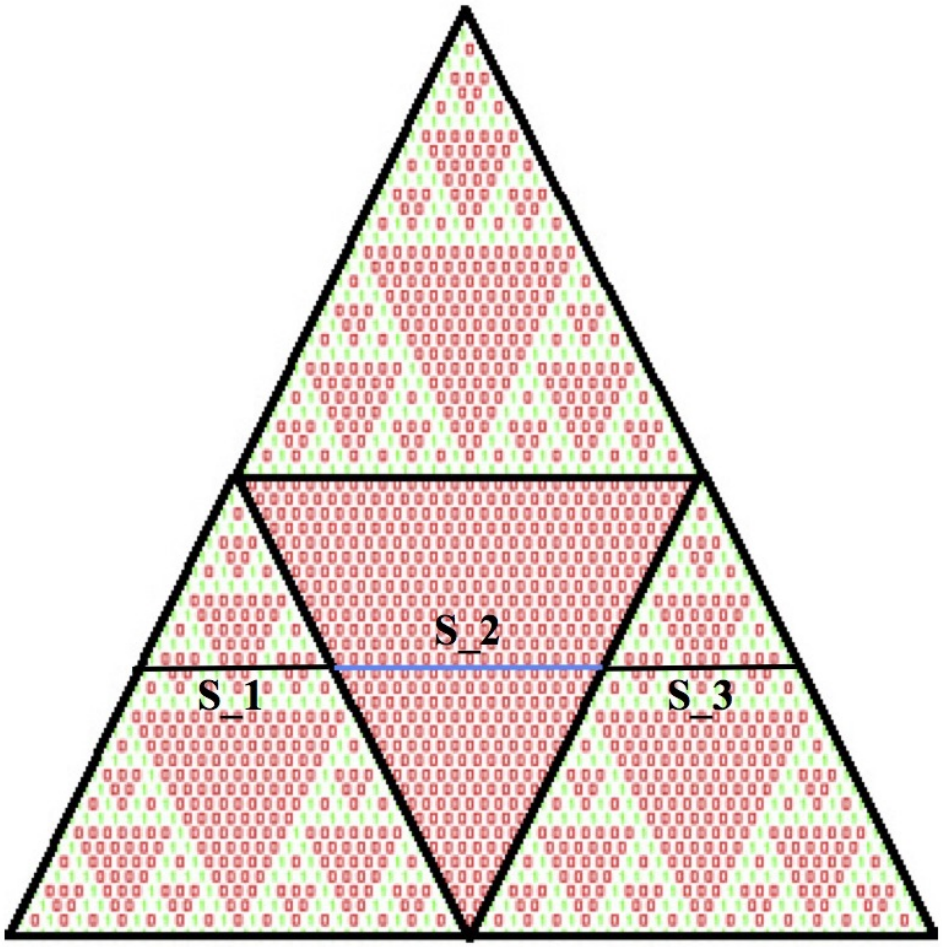


FIGURE 8. On a given induction step, we decompose a row as shown into $S_1$, $S_2$ and $S_3$.

gives us the desired result. Let $L$ be the length of $S_2$. We need only check that $f(L) = f(x_j)$. Fortunately, this follows from the same argument made before since $L = z_j$, and $x_j = z_j$ except when $z_j = 0$. Thus, the theorem is proven. □

## 6. Appendix

### 6.1. Statement of Lucas's Theorem.

**Theorem 8** (Lucas's Theorem)**.** *Let $m, k$ be integers, and let $p$ be prime. Writing $m, k$ in base $p$, we see that $m = m_n p^n + \cdots + m_1 p + m_0$ and $k = k_n p^n + \cdots + k_1 p + k_0$ for*

*some $n$. Then*

$$\binom{m}{k} \equiv \prod_{i=0}^{n} \binom{m_i}{k_i} \mod p$$

6.2. **Data for the Mod 2 Case.** The following table shows all strings of length $n$ for small $n$ which are *not* in $P_2$.

| $n=4$ | $n=5$ | $n=6$ | | | |
|---|---|---|---|---|---|
| $(1,0,1,1)$ | $(0,1,0,0,1)$ | $(0,0,1,0,0,1)$ | $(0,1,1,1,0,0)$ | $(1,0,1,1,1,0)$ | $(1,1,1,0,1,1)$ |
| $(1,1,0,1)$ | $(0,1,0,1,1)$ | $(0,0,1,0,1,1)$ | $(0,1,1,1,0,1)$ | $(1,0,1,1,1,1)$ | $(1,1,1,1,0,1)$ |
| | $(0,1,1,0,1)$ | $(0,0,1,1,0,1)$ | $(1,0,0,0,1,1)$ | $(1,1,0,0,0,1)$ | |
| | $(0,1,1,1,0)$ | $(0,0,1,1,1,0)$ | $(1,0,0,1,0,0)$ | $(1,1,0,0,1,0)$ | |
| | $(1,0,0,1,0)$ | $(0,1,0,0,1,0)$ | $(1,0,0,1,0,1)$ | $(1,1,0,1,0,0)$ | |
| | $(1,0,1,1,0)$ | $(0,1,0,0,1,1)$ | $(1,0,0,1,1,1)$ | $(1,1,0,1,0,1)$ | |
| | $(1,0,1,1,1)$ | $(0,1,0,1,1,0)$ | $(1,0,1,0,0,1)$ | $(1,1,0,1,1,0)$ | |
| | $(1,1,0,1,0)$ | $(0,1,0,1,1,1)$ | $(1,0,1,0,1,1)$ | $(1,1,0,1,1,1)$ | |
| | $(1,1,0,1,1)$ | $(0,1,1,0,1,0)$ | $(1,0,1,1,0,0)$ | $(1,1,1,0,0,1)$ | |
| | $(1,1,1,0,1)$ | $(0,1,1,0,1,1)$ | $(1,0,1,1,0,1)$ | $(1,1,1,0,1,0)$ | |

## 7. Who Did What?

As always, much of the work was a team effort. That said, Rogers wrote most of the initial code we used to test hypotheses that resulted in the Heuristic Results, and also wrote the section about zeros. Ben spent much of his time looking at the case where $p = 2$, and wrote the corresponding section as well as much of the introduction. Connor looked at more cases with prime moduli and introduced our main approach for looking at such cases, which is described in Section 3. We all contributed diagrams.