

Sums of Diagonals in Pascal's Triangle

*Jamisen McCrary**, *Russell May*



Jamisen McCrary is an undergraduate student at the University of Kentucky, studying mechanical engineering. He was previously a student in the Craft Academy at Morehead State University. His inspiration for the research topic was his prior appreciation for Pascal's Triangle and finding patterns. Apart from mathematics, Jamisen enjoys listening to music, creating art, working with electronics, programming, 3D modeling, cooking, and deep thinking. His overarching goal in life is to become a professor of engineering at a secondary education institution and build up the next generation of engineers to their fullest potential.

Russell May teaches math at Morehead State University and coaches the problem-solving club for the Craft Academy, a dual-credit academy for talented high school students in Kentucky. He is interested in combinatorics, especially problems with nice generating function solutions.



Abstract

We analyze sums of entries on diagonals of integer slope in Pascal's triangle, obtain a recurrence relation that these diagonal sums obey, and compute their generating function. We use the generating function to approximate the exponential growth of the diagonal sums.

1 Introduction

A question from a national high school math competition poses (problem 3, page 274 of [4]): how many subsets, $d(n)$, of the integers in the interval $1 \dots n + 1$ have the property that their least element coincides with their cardinality? Here is a solution: let $k \geq 1$ be the common value of the least element and cardinality of the subset. Then, besides the least element, the subset must contain a choice of the remaining $k - 1$ of the $n + 1 - k$ integers in the interval $k + 1 \dots n + 1$. So, the solution is $d(n) = \sum_{k \geq 1} \binom{n+1-k}{k-1}$.

*Corresponding author: Jamisen.McCrary@uky.edu

However, this sum can be changed to a more telling form. By the symmetry property for binomial coefficients $\binom{a}{b} = \binom{a}{a-b}$, each $d(n)$ is equivalent to $\sum_{k \geq 1} \binom{n+1-k}{n+2-2k}$. With the substitution $r = \lfloor n/2 \rfloor + 1 - k$, we can change the subtraction in the binomial coefficients to addition. So,

$$d(n) = \sum_{r \geq 0} \binom{\lfloor \frac{n}{2} \rfloor + r}{(-n) \bmod 2 + 2r}. \quad (1)$$

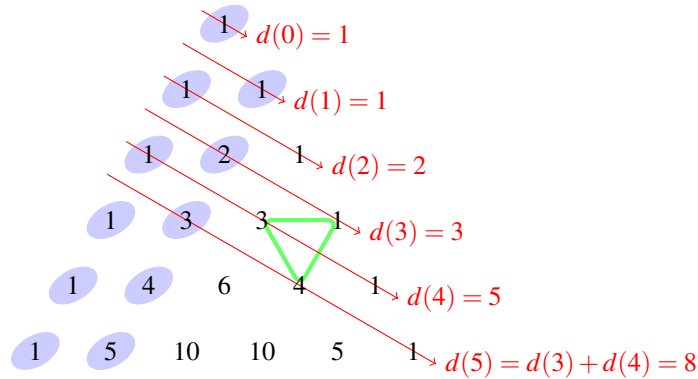


Figure 1: Sums of entries in Pascal's triangle on diagonals of slope 2 in red and intercepts in indigo. These sums coincide with the Fibonacci numbers. The triangle in green indicates how the additive identity for Pascal's triangle leads to the recurrence relation for the $d(n)$'s.

The sums in Equation 1 are depicted in Figure 1 as diagonals in Pascal's triangle, where $d(n)$ is the sum of the entries on the n^{th} diagonal from the top. These diagonals have a *slope* of 2 in the sense that if a diagonal passes through an entry in some row and position, then the diagonal also passes through the entry in the next row whose position is two more than the previous one. Each diagonal also has an *intercept*, i.e., the entry in the uppermost row with non-negative position through which the diagonal passes. The n^{th} diagonal has an intercept at $\binom{\lfloor n/2 \rfloor}{(-n) \bmod 2}$, meaning that the uppermost row that the diagonal passes through is $\lfloor n/2 \rfloor$ and the position in this row is $(-n) \bmod 2$. It is both well-known and easy to prove that the sequence $\langle d(n) : n \geq 0 \rangle$ coincides with the Fibonacci sequence beginning with two one's: 1, 1, 2, 3, 5, 8, 13, ... (see [5]). Each entry on a diagonal in Pascal's triangle is the sum of the two entries directly above it. Since these two entries lie on the two previous diagonals, then $d(n+2) = d(n) + d(n+1)$, which is the same recurrence relation that the Fibonacci numbers satisfy.

The problem from this high school competition inspires many questions. What are the sums of the entries on diagonals of slope greater than two? Are these sums on steeper diagonals also famous sequences, like the Fibonacci sequence? The terms in the Fibonacci sequence grow like powers of the golden ratio—what is the growth rate of the sums on steeper diagonals?

The general question we consider here is to determine the sum of the entries on the n^{th} diagonal of slope h in Pascal's triangle. We denote this sum by $d_h(n)$. Historically, sums

of entries on diagonals of various slopes have already been considered, for instance, in [1] and [2]. In fact, [1] even considers slopes with *rational* values and obtains recurrence relations for these sums. However, we extend their analyses by computing the generating function for the diagonal sums and then use the generating function to approximate their exponential growth.

As a brief review, we highlight a couple basic properties of Pascal's triangle. Pascal's triangle is depicted in a hexagonal lattice in a half-plane with a numerical entry in each cell. The cells in the rows are indexed by integers, and the rows are indexed by non-negative integers. The position of an entry in one row is the same as the entry to the left in the row below. The entry in the r^{th} position of the n^{th} row coincides with the binomial coefficient $\binom{n}{r}$. If $0 \leq r \leq n$, then $\binom{n}{r}$ is positive. Otherwise, it is zero for $r < 0$ or $r > n$. The additive identity asserts that $\binom{n}{r}$ is the sum of the entries in the row above to the left $\binom{n-1}{r-1}$ and right $\binom{n-1}{r}$.

To be definitive, we say a *diagonal* in Pascal's triangle is a line which passes through entries in the triangle. For a diagonal to have slope h means that if the diagonal passes through $\binom{a}{b}$ in one row, then it also passes through $\binom{a+1}{b+h}$ in the next row. We say the *intercept* of a diagonal is the uppermost row and the non-negative position in this row which the diagonal passes through. Thus, if $a \geq 0$ and $0 \leq b < h$, then $\sum_{r \geq 0} \binom{a+r}{b+hr}$ represents the sum of the entries on the diagonal in Pascal's triangle with slope h and intercept $\binom{a}{b}$. Diagonals are enumerated from the top down. So, if the diagonal through an entry has index n , then the diagonal through the entry to the left in the same row has index $n+1$, and the diagonal through the entry to the left in the next row has index $n+h$. There are h diagonals of slope h with intercepts in each row, except for the top row which just has a single intercept.

2 Diagonals of Slope Three

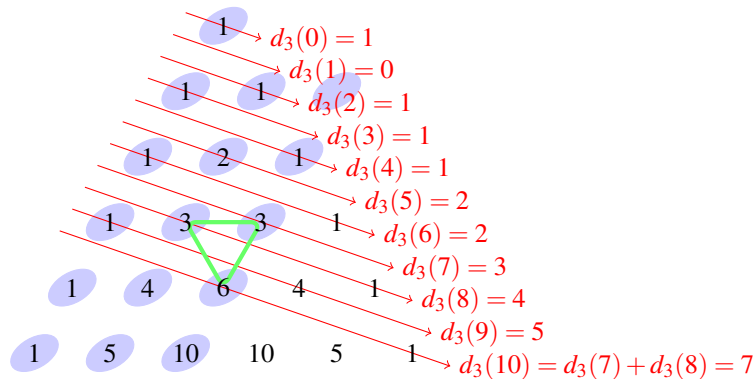


Figure 2: Pascal's triangle with sums of diagonals of slope 3 in red and intercepts in indigo. The sums of the diagonals coincide with the Padovan numbers. The triangle in green shows how the additive identity for Pascal's triangle leads to the recurrence relation for the $d_3(n)$'s

Now let's consider diagonals of slope 3 in Pascal's triangle, as depicted in Figure 2. Let $d_3(n)$ denote the sum of the entries on the n^{th} diagonal of slope 3. Note that the initial diagonal for $n = 0$ goes through the apex of Pascal's triangle, and so $d_3(0) = 1$, but the next diagonal for $n = 1$ only passes through the intercept at $\binom{1}{2}$, and so $d_3(1) = 0$. In general, the intercepts of diagonals snake through the first three positions of each row from right to left, except for the top row which only includes one position. So, the intercept of the n^{th} diagonal goes through row $\lceil n/3 \rceil$ at position $(-n) \bmod 3$. Since the slope of each diagonal is 3, the position of each entry on a diagonal is three more than the previous row. Therefore, the sum of the entries on the n^{th} diagonal of slope 3 is

$$d_3(n) = \sum_{r \geq 0} \binom{\lceil n/3 \rceil + r}{(-n) \bmod 3 + 3r}.$$

The sums of the diagonals of slope 3 coincide with the so-called Padovan sequence $\langle p_n : n \geq 0 \rangle = 1, 0, 1, 1, 1, 2, 2, 3, 4, 5, 7, 9, 12, \dots$. The Padovan sequence satisfies the recurrence relation $p_{n+3} = p_n + p_{n+1}$ for all $n \geq 0$, and the terms p_n in the Padovan sequence grow asymptotically as αr^n , where $r \approx 1.3247$ is the real root of the polynomial $x^3 - x - 1$ and $\alpha = 1/(2r + 3)$ (see [6]). The fact that the sequence $\langle d_3(n) : n \geq 0 \rangle$ of diagonals of slope 3 satisfies the same recurrence relation is a direct consequence of the additive identity for Pascal's triangle $\binom{a+1}{b} = \binom{a}{b-1} + \binom{a}{b}$. Suppose the diagonal that passes through entry $\binom{a}{b}$ has index n . Then the diagonal that passes through the entry immediately to the left $\binom{a}{b-1}$ has index $n + 1$, and the diagonal that passes through the entry $\binom{a+1}{b}$ to the left in the row below has index $n + 3$. Since the additive identity holds uniformly for all entries on these diagonals, then $d_3(n+3) = d_3(n) + d_3(n+1)$.

3 Diagonals of Integer Slope

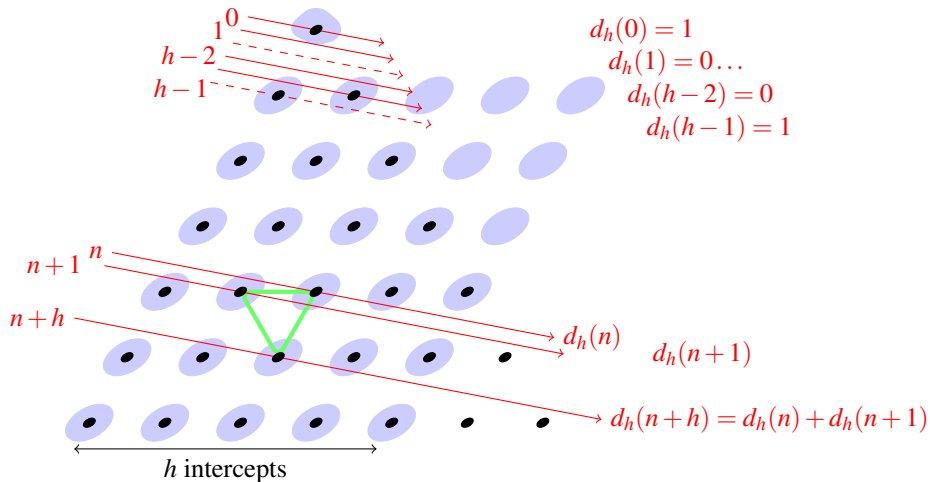


Figure 3: The first rows of Pascal's triangle with sums of diagonals of integer slope h in red and intercepts in indigo. The triangle in green shows how the additive identity for Pascal's triangle leads to the recurrence relation for the $d_h(n)$'s.

For the general case, we consider diagonals of integer slope h in Pascal's triangle, enumerated from the top down. We denote the sum of the entries on the n^{th} diagonal of slope h by $d_h(n)$. Figure 3 displays these diagonals. Note that the initial diagonal for $n = 0$ goes through the apex of Pascal's triangle, and so $d_h(0) = 1$, but the next diagonals for $n = 1 \dots h - 2$ have intercepts at $\binom{1}{r}$ in the first row with $r > 1$, and so $d_h(1) = \dots = d_h(h - 2) = 0$. The diagonal of index $h - 1$ only passes through the intercept at $\binom{1}{1}$, and so $d_h(h - 1) = 1$. Together, these give the initial conditions

$$d_h(0) = 1, \quad d_h(1) = \dots = d_h(h - 2) = 0, \quad d_h(h - 1) = 1. \quad (2)$$

The intercepts of the remaining diagonals continue to snake through the first h positions of each row from right to left. Therefore, the row of the intercept of the n^{th} diagonal is $\lceil n/h \rceil$, and its position in this row is $(-n) \bmod h$. The position of each entry in a row on the n^{th} diagonal is h more than the previous row. Therefore,

$$d_h(n) = \sum_{r \geq 0} \binom{\lceil \frac{n}{h} \rceil + r}{(-n) \bmod h + hr}.$$

This is an explicit representation of $d_h(n)$ as a sum. For the purposes of approximation, however, it is more useful to have a recursive representation. A recurrence relation for the $d_h(n)$'s is a direct consequence of the additive identity for Pascal's triangle that $\binom{a+1}{b} = \binom{a}{b-1} + \binom{a}{b}$. Suppose the n^{th} diagonal of slope h passes through entry $\binom{a}{b}$. Then, the diagonal that passes through the entry immediately to the left $\binom{a}{b-1}$ has index $n + 1$, and the diagonal that passes through the entry in the next row to the left $\binom{a+1}{b}$ has index $n + h$. Since the additive identity holds uniformly for all entries on the diagonals, then

$$d_h(n + h) = d_h(n) + d_h(n + 1), \quad \text{for all } n \geq 0. \quad (3)$$

This recurrence relation is linear, has constant coefficients, and is of degree h . We obtain an asymptotic approximation for $d_h(n)$ in the Section 5. Figure 4 shows a graph of the logarithm of $d_h(n)$ for $h = 20$ and $n = 1 \dots 1200$, based on Equation 3. This graph prominently shows damped oscillations of period h , but modulo these oscillations the graph shows simple exponential growth for the diagonal sums.

4 Generating Function

An often-used tool for analyzing combinatorial sequences is the generating function. The ordinary generating function of the sequence $\langle a_n : n \geq 0 \rangle$ is $\sum_{n \geq 0} a_n x^n$. It can be thought of as a formal power series or, wherever it converges, a function of complex numbers. Wilf in section 1.2 of [3] gives a five-step method for converting a recurrence relation describing a sequence to its generating function: clarify the set of valid values of the free variable in the recurrence relation, name the generating function, multiply each instance of the recurrence by an appropriate power of the variable of the generating function and sum over the valid values, express both sides of the resulting equation in terms of the generating function, and finally solve the resulting equation for the generating function. The initial conditions in Equation 2 give the

values of $d_h(n)$ for $n = 0 \dots h-1$, and the general recurrence relation in Equation 3 determines all the rest of the values for $n \geq h$. We define the generating function $D_h(x) = \sum_{n \geq 0} d_h(n)x^n$. Multiplying each term in the initial conditions of Equation 2 and recurrence relation of Equation 3 by the corresponding power of x and summing, we get $(D_h(x) - 1 - x^{h-1})/x^h = D_h(x) + (D_h(x) - 1)/x$. Solving for the generating function results in an amazingly simple expression:

$$D_h(x) = \frac{1}{1 - x^{h-1} - x^h}. \quad (4)$$

5 Approximation of Sums

The generating function in Equation 4 for the sequence $\langle d_h(n) : n \geq 0 \rangle$ is first and foremost a rational function, i.e., a ratio of polynomials. In this case the rational function has a numerator $f(x) = 1$ and denominator $g(x) = 1 - x^{h-1} - x^h$. Using basic tools of calculus, it is straightforward to approximate the exponential growth of any sequence whose generating function is rational. The first step is to determine the partial fraction decomposition of the generating function. Let R_h denote the collection of roots of the denominator $g(x)$. Since $g(x)$ and its derivative $g'(x) = -x^{h-2}(h-1+hx)$ have no common roots, then none of the roots of $g(x)$ are repeated. So, the partial fraction decomposition of the generating function has the form

$$D_h(x) = \sum_{r \in R_h} \frac{a_r}{x-r},$$

where $a_r = \frac{f(r)}{g'(r)}$. Each term in the partial fraction decomposition represents a geometric series, as follows:

$$\frac{a_r}{x-r} = -\frac{a_r}{r(1-\frac{x}{r})} = -\frac{a_r}{r} \sum_{m \geq 0} \frac{x^m}{r^m}$$

We can extract the coefficient of each term of the decomposition with the coefficient extraction operator. By definition, if $f(x) = \sum_{n \geq 0} a_n x^n$, then $[x^n]f(x) = a_n$. Then,

$$\begin{aligned} & [x^n]D_h(x) \\ &= \sum_{r \in R_h} [x^n] \sum_{m \geq 0} -\frac{a_r}{r} \frac{x^m}{r^m} \\ &= \sum_{r \in R_h} -\frac{f(r)}{r g'(r)} r^{-n}. \end{aligned} \quad (5)$$

The terms with the largest contribution to $d_h(n)$ in Equation 5 are the ones whose roots have the smallest modulus. In this case, we will shortly see that there is a single real root \tilde{r} of $g(x)$ with the smallest modulus, and this root is called the *dominant singularity* of the generating function. The exponential growth approximation for $d_h(n)$ concentrates solely on this singularity, that is,

$$d_h(n) \approx -\frac{f(\tilde{r})}{\tilde{r} g'(\tilde{r})} \tilde{r}^{-n}. \quad (6)$$

To find the dominant singularity, first consider the case when the slope h of the diagonals is even. When $g(x) = 1 - x^{h-1} - x^h$ is graphed on the real line, we see that $g(-\infty) = g(\infty) = -\infty$, $g(-1) = g(0) = 1$, and $g(1) = -1$. Because of the sign

changes, there are real roots in the intervals $(-\infty, -1)$ and $(0, 1)$. Since the derivative $g'(x) = -x^{h-2}(h-1+hx) > 0$ iff $x < -\frac{h-1}{h}$, then these are the only real roots. To see that \tilde{r} has a smaller modulus than any of the complex roots, note that if $|z| < \tilde{r}$, then $|z^h + z^{h-1}| \leq |z^h| + |z^{h-1}| < \tilde{r}^h + \tilde{r}^{h-1} = 1$, excluding the possibility that such z 's could be a root of g . So, the dominant singularity \tilde{r} of this generating function must be the real root in the interval $(0, 1)$. By a similar analysis, when h is odd, the dominant singularity is still in $(0, 1)$. For $h \geq 5$, it is impossible to express \tilde{r} in terms of radicals. However, it is easy to approximate for large values of h . If h is large, then $h-1$ and h are both nearly equal to $h-1/2$, and so $g(x) \approx 1 - 2x^{h-1/2}$. Therefore, the dominant singularity is approximately $\tilde{r} \approx 2^{-1/(h-1/2)}$. At this value, $f(\tilde{r}) = 1$ and $g'(\tilde{r}) \approx -h + 1/2$. Plugging in these values into Equation 6 results in

$$d_h(n) \approx \frac{1}{h-1/2} 2^{\frac{n+1}{h-1/2}} \quad (7)$$

The graph of the logarithm of this approximation appears as a line in Figure 4 and shows close agreement to the graph of the logarithm of $d_h(n)$. However, it is inherent that since only an approximation was used for the dominant singularity, then this approximating line must eventually diverge from the exact values. Of course, the exponential growth in Equation 7 does not account for the oscillations that are prominent in the graph of $d_h(n)$ in Figure 4. These oscillations are the result of the complex roots in the partial fractions decomposition of the generating function and will be the object of further study.

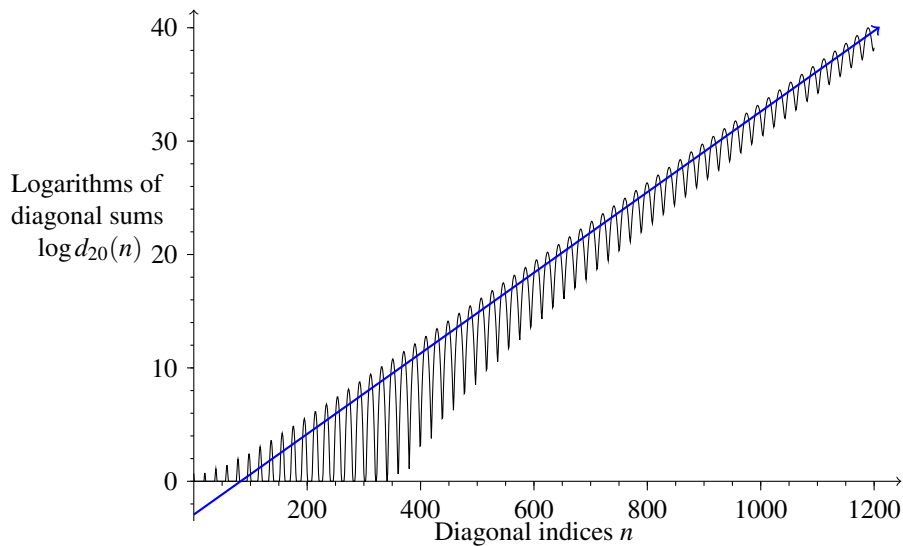


Figure 4: A graph of the logarithms of the sums of the entries on the first 1200 diagonals with slope $h = 20$ in Pascal's triangle, along with the linear approximation from Equation 7.

Bibliography

- [1] Green, Thomas M. Recurrent Sequences and Pascal's Triangle. *Mathematics Magazine*, 41(1), 13-21, (1968). <http://www.jstor.org/stable/2687953>
- [2] Joseph A. Raab. A Generalization of the Connection Between the Fibonacci Sequence and Pascal's Triangle. *The Fibonacci Quarterly*, 3(3), 21-31, (1963) <https://www.fq.math.ca/Scanned/1-3/raab.pdf>
- [3] Herbert S. Wilf. *Generatingfunctionology*. A K Peters/CRC Press, (2005).
- [4] Paul J. Karafiol et al. American Regions Math League Power & Local Contests 2009–2014. ARML, (2015).
- [5] N. J. A. Sloane. Online Encyclopedia of Integer Sequences. <http://oeis.org/A000045>
- [6] N. J. A. Sloane. Online Encyclopedia of Integer Sequences. <http://oeis.org/A000045>