MIT PRIMES CIRCLE 2024

QUINTEN JIN, EMRE KOCAMAN

Abstract. This paper explores the intricate world of combinatorics, focusing on several key principles and their applications in solving complex counting problems. Through the usage of the Principle of Inclusion and Exclusion (PIE), Catalan numbers, and generating functions, we explore a variety of problems ranging from language overlaps in multilingual classes to domino tiling puzzles. The initial section uses PIE to dissect a classroom language scenario, providing a foundational understanding of counting without overestimation. Further, we investigate Catalan numbers to elucidate counting structures maintaining non-negative cumulative sums and path constraints. The paper also tackles the Domino Tiling Problem, introducing a recursive relationship akin to the Fibonacci sequence to determine tiling configurations for larger boards. Lastly, we expand on generating functions, a powerful tool for encoding sequences and solving combination problems systematically.

1. Introduction

Combinatorics is the study of counting. Both counting which we learn when we are 3 and finding the generating recurrences fall under the same study. People have always found it necessary so this study was created the moment humanity existed. Combinatorics isn't useful just in harsh math exams, it is useful everywhere from trying to crack down the password of your brother's i-pad to finding the number of alternative routes to take from one place to another.

In this paper, we will start with PIE which stands for the Principle of Inclusion and Exclusion. We will first use PIE in basic language problems. Then, we will use it in harder problems which seem irrelevant to PIE but actually is PIE. Next, we will continue with special numbers known as Catalan Numbers. These numbers are special numbers which can be used in counting and in specific problems. We will continue with Closed Forms. We will learn how to create explicit formulas from recursive formulas. We will prove the Fibonacci explicit formula. After that, we will learn about Recursions. We will create recursive formulas for complicated problems. We will conclude with Generating Functions and Choose Function for other complicated problems.

2. The Principle of Inclusion-Exclusion

Theorem 2.1 (PIE). *PIE stands for Principle of Inclusion and Exclusion. This easy method is used to avoid over counting and over subtracting. It's generalized formula is:*

$$\left|\bigcup_{i=1}^{n} S_k\right| = \sum_{I\subseteq[n]} (-1)^{(|I|+1)} \left|\bigcap_{i\in l} S_i\right|$$

The left side of the theorem is $S_1 \cup S_2 \cup S_3 \cdots \cup S_k$. The right side indicates that if the number of intersecting sets are even, they are subtracted. If the number of intersecting sets are odd, they are added.

The following equality proves PIE:

$$0 = \binom{k}{0} - \binom{k}{1} + \binom{k}{2} - \binom{k}{3} + \dots,$$

quintenjin@gmail.com, eakocaman@gmail.com.

A set should have a value of 1 if it includes a member and should have a value of 0 if it doesn't. k is the number of sets which have the value 1. k doesn't have to be equal to the number of all the sets. This equality is rearranged as $1 = \binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots,$*. In PIE, all the sets are added, the two-intersecting sets are subtracted, the three-intersecting sets are added and it goes on and on. The number of sets are* $\binom{k}{1}$*, the number of two intersecting sets are* $\binom{k}{2}$ *and the number of three intersecting sets are* $\binom{k}{3}$*. PIE avoids over counting and over subtracting. The set which contains all the sets have the value 1. The* $\binom{k}{2}$ *is subtracted from* $\binom{k}{1}$ *and* $\binom{k}{3}$ *is added. It goes on and on.*

Let's understand PIE better in some example problems.

Problem. *30 people speak French and 30 people speak German in a classroom. 10 people speak both. Everyone speaks 1 language at least. How many people are there?*

In this problem, when French speakers are added to German speakers, there are 60 people. However, the people speaking both were added twice. Thus, the people who speak both should be subtracted from 60 people. Bingo! The answer is 50. Before moving on, this problem could be written differently: $\mid A \cup B \mid = \mid A \mid + \mid B \mid - \mid A \cap B \mid$. Set $\mid A \mid$ represents the French speakers and Set $\mid B \mid$ represents the German speakers.

Problem. *In a classroom, 20 people play volleyball, 10 people play basketball and 20 people play baseball. 3 people play both baseball and volleyball, 6 people play volleyball and baseball and 4 people play baseball and basketball. 3 people play all of them. Everyone at least plays 1 sport. How many people are there in this classroom?*

In this problem, when volleyball players, basketball and baseball players are added, we get 50. However, we over counted students who play multiple sports so we have to subtract the ones who play two sports. 13 people play at least two sports so we have to subtract them. The answer is not 37 though. We over subtracted the ones who play 3 sports. Thus, we have to add them. The answer is 40. This problem can also be written differently: $\mid A \cup B \cup C \mid \; = \mid A \mid + \mid B \mid + \mid C \mid - \mid A \cap B \mid - \mid A \cap C \mid - \mid B \cap C \mid + \mid A \cap B \cap C \mid$, where $\mid A \mid$ represents basketball, $\mid B \mid$ represents baseball and $\mid C \mid$ represents volleyball.

Problem. *We want to rearrange the word "ABRACADABRA" such that there are no "ABRA"s. How many ways are there?*

Before taking any steps, we should find out the number of all arrangements. It is $\frac{11!}{5!2!2!}$. Now, we have to subtract the number of arrangements which don't satisfy what we want. In order to do that, let's treat "ABRA" as one object. Then, the number of unsatisfied arrangements could be found as $\frac{8!}{3!}$. When we subtract $\frac{8!}{3!}$ from $\frac{11!}{5!2!2!}$, we should be good, shouldn't we? But wait a second. Aren't we over subtracting the arrangements in which there are two "ABRA"'s? Thus, we should add the number of arrangements which have two "ABRA"'s. As a result, our answer should be $\frac{11!}{5!2!2!} - \frac{8!}{3!} + \frac{5!}{2!}$

Problem. $N = 2 \times 3 \times 5 \times 7 \times 11 \times 13$*. How many numbers smaller than N don't have a common divisor with N other than 1?*

This problem doesn't seem related to PIE, does it? Well, PIE is the exact thing we are going to use. Before moving on, the number of all alternatives is N. We want to subtract the number of integers which are divisible by 2,3,5,7,11 and 13. We should subtract $\frac{N}{2}$, $\frac{N}{3}$, $\frac{N}{5}$, $\frac{N}{7}$, $\frac{N}{11}$ and $\frac{N}{13}$.However, we are over subtracting the numbers which are divisible by 2 prime factors. We should add them. We are over counting the numbers divisible by 3 prime numbers. Thus, we should subtract them. Overall, we are using PIE. In algebraically, our answer should be written in $N - (\frac{N}{2} + \frac{N}{3} + \frac{N}{5} + \frac{N}{7} + \frac{N}{11} + \frac{N}{13}) + (\frac{N}{2\times3} + \frac{N}{2\times5} \ldots) - (\frac{N}{2\times3\times5} - \ldots) + \frac{N}{2\times3\times5\times7\times11\times13} = N(1-\frac{1}{2}) \times (1-\frac{1}{3}) \times (1-\frac{1}{5}) \times (1-\frac{1}{7}) \times (1-\frac{1}{11}) \times (1-\frac{1}{13}) = N(\frac{1}{2}) \times (\frac{2}{3}) \times (\frac{4}{5}) \times (\frac{6}{7}) \times (\frac{10}{11}) \times (\frac{12}{13}) = 2 \times 4 \times 6 \times 10 \times 12.$

3. Catalan numbers

Catalan numbers are special numbers which are very useful in some problem styles.

Definition 3.1 (Catalan numbers)**.**

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

Let's say there are n (+1)'s and n (-1)'s. C_n is the number of arrangements in which when numbers are formed, the sum of numbers is always zero or bigger when the numbers are added one by one from left to right.

For example, let's say n is 2.
1.(+1),(+1),(-1),(-1) is okay.
2.(+1),(-1),(+1),(-1) is okay.
3.(-1),(-1),(+1),(+1) is NOT okay.
4.(+1),(-1),(-1), (+1) is NOT okay.

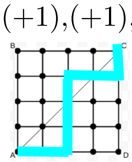
In order to find the number of these arrangements which satisfy this, we have to find the number of arrangements which don't and subtract it from the total number of possibilities. The total number of possibilities are $\binom{2n}{n}$. Now the tricky part is how to find the number of possibilities which don't satisfy what we want. In those possibilities, when the first time the number of -1's are greater, that -1 has to be the 2k+1 th number. Except the 2k+1 th number, when every number's sign is changed, there are going to be n+1 -1's and n-1 +1's. The number of the arrangements of n+1 (-1)'s and n-1 (+1)'s are $\binom{2n}{n+1}$. When we subtract it from $\binom{2n}{n}$, we get $\binom{2n}{n} - \binom{2n}{n+1} = \binom{2n}{n} - \frac{n}{n+1}\binom{2n}{n} = \frac{1}{n+1}\binom{2n}{n}$.The formula is therefore $C_n = \frac{1}{n+1}\binom{2n}{n}$.

Problem. *How many ways are there to arrange 6 open parentheses and 6 closed parentheses so that all the parentheses are paired up?*

This problem can obviously be solved with Catalan numbers. If (is considered as +1 and) is considered as -1, this problem can be solved easily. Meaning that, for example,((()()) is same as (+1),(+1),(+1),(-1),(+1),(-1),(-1) and))()((is same as (-1),(-1),(+1),(-1),(+1),(+1). The first scenario is okay and the second one isn't. Thus, we see that the answer is equal to C_6. All that has to be done is plugging 6 to the n value in the closed formula. The answer should be 95040.

Problem. *A bunny wants to go to the opposite corner of a 4 times 4 square. That bunny can step on diagonal but can't pass the diagonal. How many ways can that bunny go to the opposite corner?*

This problem can be solved with Catalan numbers as well. In Catalan numbers, the number of (-1)'s was never larger than the number of (+1)'s while adding them from left to right one by one. In 4 times 4 square, if the number of vertical moves is more than the number of horizontal moves any time, the bunny would pass the diagonal. Meaning that, in order not to cross the diagonal, the number of vertical lines should be similar to (-1) and horizontal lines should be considered as (+1). For example, (+1),(+1),(-1),(-1),(-1),(+1),(+1),(-1) is same as:



As a result, n should be replaced with 4 on the formula. The answer should be 336.

Theorem 3.2.

$$C_n = C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-2}C_1 + C_{n-1}C_0$$

Why is this theorem true? Let's look at the parentheses problem again. In the parenthesis problem, let's say that these parenthesis are separated into two groups. One group has k pairs and the other has $n-k$ pairs. Those k pairs could be arranged in C_k ways and $n-k$ pairs could be arranged in C_{n-k} ways. Thus, when those pairs are separated as k and n-k, the number of possible arrangements are $C_k C_{n-k}$.

The sum of all of the possible divisions could be found by putting all the values to k from 0 to $n-1$ and then summing all of them.

4. Closed Forms

In this section, we are going to talk about how to create explicit formulas for closed forms. Geometric sequences are going to be used in order to create explicit formulas. We'll understand this better in example problems.

Problem. *$A_n = 4 \times A_{n-1} - 3 \times A_{n-2}$. Initial term is $A_0 = 5$ and $A_1 = 13$. What is the explicit formula?*

Let's consider that a geometric sequence also satisfies this closed form. Therefore, we can write this equation: $x^n = 4x^{n-1} - 3x^{n-2}$. That term is rearranged as $(x-1) \times (x-3) = 0$. x has two solutions. We are going to combine these solutions algebraically so our explicit formula should be $A_n = a(3)^n + b(1)^n$. Since $A_0 = 5$ and $A_1 = 13$, when we replace n with 0, we should get the equation $b + a = 5$ and when we replace n with 1, we get the equation $b + 3a = 13$. When these equations are solved, b is equal to 1 and a is equal to 4. The explicit formula for this closed formula is thus $A_n = 4(3)^n + 1$.

Theorem 4.1.

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

Proof. Fibonacci is a recurrence in which F_n is found by adding the previous two terms. Now, we are going to prove its explicit formula with our principle. The closed form of Fibonacci is $F_n = F_{n-1} + F_{n-2}$. Let's say that a geometric sequence satisfies this formula. Thus, we write like $x^n = x^{n-1} + x^{n-2}$. When this is rearranged, it can be seen that $x^2 - x - 1 = 0$. The x values which satisfy this equation is $\frac{1+\sqrt{5}}{2}$ and $\frac{1-\sqrt{5}}{2}$. $\varphi = \frac{1+\sqrt{5}}{2}$ and $\overline{\varphi} = \frac{1-\sqrt{5}}{2}$. We should combine φ and $\overline{\varphi}$ algebraically. Therefore, our explicit formula is $F_n = A\varphi^n + B\overline{\varphi}^n$. Since $F_0 = 0$ and $F_1 = 1$, these equations should be set up: $A + B = 0$ and $A(\varphi) + B(\overline{\varphi}) = 1$. When these equations are solved, $A = \frac{1}{\sqrt{5}}$ and $B = \frac{-1}{\sqrt{5}}$. Therefore, $F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$. □

Problem. *$B_n = 4 \times B_{n-1} - 4 \times B_{n-2}$. $B_0 = 2$ and $B_1 = 4$ What is B_{100}?*

Like we've done before, let's say a geometric sequence satisfies this. We re-write it as: $x^n = 4 \times x^{n-1} - 4 \times x^{n-2}$. When we divide both sides by x^{n-2}, we get $x^2 = 4x - 4$. When we re-arrange it as $x^2 - 4x + 4 = 0$, we find $x = 2$. When we set the equations as $2 = A(2)^0$ and $4 = A(2)^1$. A is 2. As a result, $B_n = 2(2)^n$. Thus, B_{100} is 2^{101}.

5. Recursion

5.1. **Problem Statement.** The problem is to find the number of ways to cover a 2×100 board with 1×2 dominos. We denote this function by $F(n)$, where n represents the width of the board.

5.1.1. *First thoughts.* We can explore a recursive solution by starting with an example of a 2×2 square. We can cover a 2×2 square in two ways:

- **Case 1: Horizontal Domino.** If the first domino is placed horizontally, it covers two cells in the first row. Therefore, another horizontal domino must be placed directly beneath it to cover the remaining two cells. This leaves us with a $2 \times (n-2)$ board. The number of ways to complete this remaining board is given by $F(n-2)$.

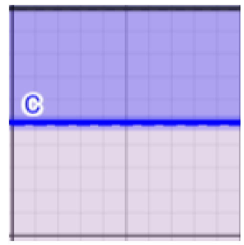


Figure 1. A horizontally placed domino on a 2x2 square

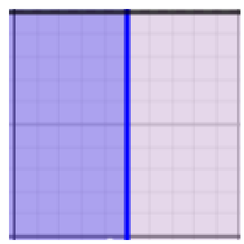


Figure 2. A horizontally placed domino on a 2x2 square

- **Case 2: Vertical Domino.** If the first domino is placed vertically, it covers one cell in each row. This leaves us with a $2 \times (n-1)$ board. The number of ways to complete this remaining board is given by $F(n-1)$.

5.1.2. *Recurrence Relation.* Given these two cases, the total number of ways to cover a $2 \times n$ board can be expressed as:

$$F(n) = F(n-1) + F(n-2)$$

For this recursion to work, we need to define initial conditions:

- $F(0) = 1$: There's one way to cover a 2×0 board — by doing nothing.
- $F(1) = 1$: There's only one way to cover a 2×1 board — by placing one vertical domino.

So if we follow the formula out, it will turn out to be

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \cdots$$

Here is another problem that implement the same technique to solve a similar problem.

5.2. **Problem Description.** The platform is to be filled with L-shaped blocks, each occupying 1x2 cells in both rows, and individual 1x1 blocks. The objective is to determine the total number of ways to fill the platform by calculating $F(n)$ for a given n.

5.2.1. *Recurrence Relation.* The placement of blocks can be categorized into two main cases:

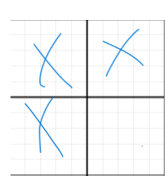


Figure 3. Case 1

Case 1.

- An L-shaped block is placed, and the space immediately following it is filled with another L-shaped block or a 1x1 block. This contributes to:
  - $F(n-2)$ when the subsequent space is filled with a 1x1 block.
  - $F(n-3)$ when another L-shaped block is used.

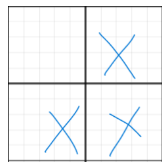


FIGURE 4. Case 2

Case 2.

- An L-shaped block is placed such that only a 1x1 block can follow, contributing to:
  - $F(n-2)$ for filling the remaining space with a 1x1 block.

Combining these cases, we derive the recurrence relation:

$$F(n) = 4F(n-2) + 2F(n-3)$$

5.2.2. *Initial Conditions.* To use the recurrence relation effectively, initial conditions are specified as:

$$\begin{aligned} F(0) &= 1 \quad \text{(the empty platform)} \\ F(1) &= 0 \quad \text{(impossible to fill a 2x1 platform)} \\ F(2) &= 2 \quad \text{(filled either by two 1x1 blocks or one 2x1 L-shaped block)} \end{aligned}$$

5.2.3. *Conclusion.* Using the established recurrence relation and initial conditions, one can compute $F(n)$ for any positive integer n, effectively determining the number of ways to fill a 2xN platform with the specified blocks.

6. Generating Functions

Picking Numbers from Two Boxes. Two boxes contain numbers as follows:

- Box 1: $\{1, 2, 3\}$
- Box 2: $\{2, 3, 4\}$

and you randomly pick one number from each box.

1(a) List all combinations and their sums. The sums obtained by choosing one number from each box are:

$$1 + 2 = 3$$
$$1 + 3 = 4$$
$$1 + 4 = 5$$
$$2 + 2 = 4$$
$$2 + 3 = 5$$
$$2 + 4 = 6$$
$$3 + 2 = 5$$
$$3 + 3 = 6$$
$$3 + 4 = 7$$

Thus, the distribution of sums is: $[3, 4, 4, 5, 5, 5, 6, 6, 7]$.

The and frequency of each sum is as follows:

- 3 appears once.
- 4 appears twice.
- 5 appears three times.
- 6 appears twice.
- 7 appears once.

1(b) Generating function representation. *Introduction to generating functions:* A generating function is a formal power series that encodes a sequence $a_0, a_1, a_2, \ldots$ as follows:

$$G(x) = \sum_{n=0}^{\infty} a_n x^n$$

The coefficient a_n represents the nth term in the sequence. Generating functions can simplify the manipulation and study of sequences, allowing us to analyze them through algebraic means.

Generating functions for the same problem: The process of choosing numbers from each box can be described with generating functions.

The generating function for choosing numbers from Box 1 is:

$$G_1(x) = x^1 + x^2 + x^3$$

This function encodes the set of numbers $\{1, 2, 3\}$ by mapping them to powers of x.

The generating function for choosing numbers from Box 2 is:

$$G_2(x) = x^2 + x^3 + x^4$$

This function similarly encodes the set of numbers $\{2, 3, 4\}$.

Multiplicative property of generating functions: When we want to combine two sequences (in this case, choosing a number from each box and summing them), we can represent this operation by multiplying their generating functions:

$$G(x) = G_1(x) \cdot G_2(x)$$

Simplifying this product:

$$G(x) = (x^1 + x^2 + x^3) \cdot (x^2 + x^3 + x^4)$$
$$G(x) = x^3 + 2x^4 + 3x^5 + 2x^6 + x^7$$

The coefficients of this polynomial correspond to the frequencies of each sum:

- x^3 has a coefficient of 1, representing one way to obtain a sum of 3.
- x^4 has a coefficient of 2, representing two ways to obtain a sum of 4.
- x^5 has a coefficient of 3, representing three ways to obtain a sum of 5.
- x^6 has a coefficient of 2, representing two ways to obtain a sum of 6.
- x^7 has a coefficient of 1, representing one way to obtain a sum of 7.

Why generating functions are useful: Generating functions are useful because:

- They encode an entire sequence into a single function.
- By treating generating functions as polynomials or power series, we can analyze and manipulate them using algebraic operations such as addition, multiplication, differentiation, and integration.
- Multiplying generating functions can show how two sequences relate to each other, as in the case of combining Box 1 and Box 2.
- Generating functions simplify the counting of combinations, permutations, and partitions, among other combinatorial structures.

6.1. **Here is another example for the use case of Generating function.**

6.1.1. *Problem Formulation.* Consider two dice, each labeled with numbers $\{1, 2, 3, 4\}$. When these dice are rolled, the possible outcomes of their sums range from 2 to 8. We aim to explore whether different sets of numbers on two dice could yield the same sum distribution.

6.1.2. *Generating Functions Approach.* The generating function for each die is given by:

$$G(x) = x^1 + x^2 + x^3 + x^4$$

The generating function for the sum of the dice, when both are rolled, is the product of the individual generating functions:

$$G(x) \cdot G(x) = (x + x^2 + x^3 + x^4)^2$$

Simplifying this, we have:

$$G(x) = (x + x^2 + x^3 + x^4)^2 = (x(1 + x + x^2 + x^3))^2$$

This expression can be rearranged by factoring to:

$$G(x) = (x(1 + x)(1 + x^2))^2 = (x^2 + x)^2(1 + x^2)^2$$

Expanding these terms, we find:

$$(x^2 + x)^2 = x^4 + 2x^3 + x^2$$

$$(1 + x^2)^2 = 1 + 2x^2 + x^4$$

Thus, the expanded product is:

$$G(x) = (x^4 + 2x^3 + x^2)(1 + 2x^2 + x^4)$$

6.1.3. *Interpretation and Alternative Dice Sets.* The coefficients in the expanded form of $G(x)$ represent the number of ways each possible sum can occur. If we seek an alternative set of dice that yields the same sum distribution, we must ensure the generating function of the new set results in the same coefficients for corresponding powers of x. Through further algebraic manipulations and substitutions, different configurations can be explored to achieve this equivalence for most of the similar cases.

Acknowledgement

Sections 2-4 were written by Emre Kocaman and Sections 5-6 were written by Quinten Jin. Authors thank Sean J Li for the invaluable mentorship he has provided throughout the program. Authors are also truly grateful for the guidance and support from Mary C Stelow and Marisa R Gaetz.