

A bouquet of counting problems and word combinatorics

1 Election \pm problem

Consider a closely fought election in West Bengal in April 2026 between two major political alliances, say Party A and Party B . As the counting of votes progresses constituency by constituency, television channels continuously display the running totals. Sometimes one party leads early, then the other catches up, and the lead may fluctuate dramatically throughout the day.

Now suppose Party A eventually wins by a very small margin: it receives $n + 1$ constituencies while Party B receives n . A natural combinatorial question arises: In how many possible counting sequences does Party A remain always ahead of (or at least tied with) Party B during the entire counting process?

Example. For $n = 2$, there are five such sequences:

$AAABB$, $AABAB$, $AABBA$, $ABAAB$, $ABABA$.

Exercise 1.1:

Find out the number of such sequences for $n = 3, 4, 5$.

2 Well-formed formulas

Suppose p and q are variables and $*$ is a binary operation (you can think of addition or multiplication, for example). Here are some logical syntactic rules to form new expressions using the symbols $p, q, *$ and parentheses—the so-called *well-formed formulas*:

1. Every variable is a well-formed formula.
2. If s and t are well-formed formulas, then so is $(s * t)$.

Example.

1. Some examples of well-formed formulas: $p, q, (p * q), ((p * p) * q)$ etc.
2. Some non-examples of well-formed formulas: $(p * p) * q, (s * t, p * q)$ etc. (In each case, think about *why* the expression is not a well-formed formula.)

Quick tests to determine if a given string of symbols is not a well-formed formula are the following.

Exercise 2.1:

1. The number of left parentheses '(' in a well-formed formula equals the number of right parentheses ')'.
2. (**Strong prefix property**) If s is a well-formed formula that is not a variable, and t is a proper left substring of s containing at least one symbol, then the number of left parentheses in t is strictly greater than the number of right parentheses in it.

Suppose you are tasked with arranging n pairs of parentheses while satisfying the **weak prefix property**: the number of left parentheses in any proper left substring is at least the number of right parentheses in it. For $n = 1$, you have only one choice, namely $()$ but for $n = 2$, you have two choices, namely $()()$ or $(())$.

Example. The following arrangements of parentheses satisfy the weak prefix property but not the strong prefix property: $(())()$, $()(())$.

Exercise 2.2:

Find all expressions for $n = 3, 4, 5$ satisfying the weak prefix property. Can you relate this problem to Exercise 1.16?

Exercise 2.3:

Let C_n denote the number you counted in the above exercise. Is it possible to compute C_{n+1} if you know C_0, C_1, \dots, C_n ?

3 Building Mountains

Did you know that the tallest mountain in the world is Mauna Kea in Hawaii, and not Mount Everest, when measured from its base to peak?

Suppose you want to build a mountain range using n up-strokes (/) and n down-strokes (\) but you can never go below the sea level.

Example. For $n = 1$, you have only one choice: $/\backslash$.

For $n = 3$, the landscape is more varied. The following are the valid mountain ranges with 3 up-strokes and 3 down-strokes: $///\backslash\backslash$, $//\backslash\backslash\backslash$, $///\backslash\backslash$, $/\backslash\backslash\backslash$, and $/\backslash\backslash\backslash$.

Exercise 3.1:

List and count the distinct mountain ranges that you can build for $n = 5$.

You might want to think about parenthesis arrangements satisfying the weak prefix property when constructing mountain ranges.

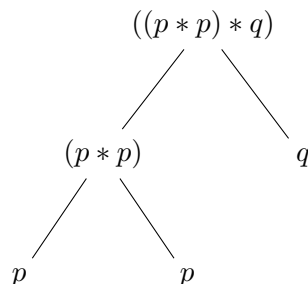
Counting mountain ranges using reflection (André's Method)

Let M_n denote the number of mountain ranges constructed using n -many of / and n -many \. The counting of M_n proceeds as follows: we will subtract the number of bad mountain ranges from the number of all possible mountain ranges, which can go below the sea level as well. Before we begin, let us look at a small example to illustrate the removal of bad mountain ranges:

5 Full Binary Trees

Full binary trees are used by computers in the evaluation of well-formed formulas from Section 2—they are rooted trees where every node has either zero or two children. A node with zero children is called a *leaf node* and a node that is not a child of any other node is called the *root node*. An *internal node* is a node that is not a leaf node.

The following picture shows how the evaluation of $((p * p) * q)$ works.



Exercise 5.1:

Count the number of full binary trees with n internal nodes for $n = 3, 4$. Is this problem related to Exercise 2.2?

Exercise 5.2:

Is it possible to view the internal nodes of a full binary tree as triangles in a triangulation of an appropriate regular polygon? (The root could be the unique triangle containing the edge 12 in a given triangulation of the polygon.)

6 Bonus: Generating functions

Recall the numbers C_n from Exercise 2.3. Define the *generating function* of this sequence to be the expression $F(x) := \sum_{n \geq 0} C_n x^n$.

Example. A simple example of a generating function is that of the *geometric series*:

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}.$$

Simple manipulations of such expressions will actually turn out to be an easy way to evaluate sequences that satisfy certain recurrence relations. Here are some rules that will help us.

Exercise 6.1:

Let $A(x) = \sum_{n \geq 0} a_n x^n$. Fix a positive integer m .

1. Express $\sum_{n \geq m} a_{n-m} x^n$ in terms of $A(x)$.
2. Express $\sum_{n \geq 0} a_{n+m} x^n$ in terms of $A(x)$.

Exercise 6.2:

- Let $A(x) = \sum_{n \geq 0} a_n x^n$ and $B(x) = \sum_{n \geq 0} b_n x^n$. Express $D(x) := \sum_{n \geq 0} (\sum_{k=0}^n a_k b_{n-k}) x^n$ in terms of $A(x)$ and $B(x)$.
- Let $A_1(x), \dots, A_m(x)$ be defined as $A_i(x) := \sum_{n \geq 0} a_{i,n} x^n$. Express $D(x) := \sum_{n \geq 0} (\sum_{k_1 + \dots + k_m = n} a_{1,k_1} a_{2,k_2} \dots a_{m,k_m}) x^n$ in terms of $A_1(x), \dots, A_m(x)$.

Exercise 6.3:

Use the recurrence relation obtained in Exercise 2.3 and the rules for manipulating generating functions that we saw above, show that $F(x) = 1 + xF(x)^2$. Use this to find an expression for $F(x)$ in terms of x .

7 Christoffel words

Given a positive integer n , a *Dyck path* (named after the German mathematician von Dyck, read *fon Daeek*) is a step-path from $(0,0)$ to (n,n) that does not cross the line segment joining the end points. Dyck paths could be read as words over the alphabet $\{x, y\}$. For $n = 3$, there are 5 Dyck paths: $xxxyyy, xyxyxy, xyxyxy, xyxyxy, xxyyxy$.

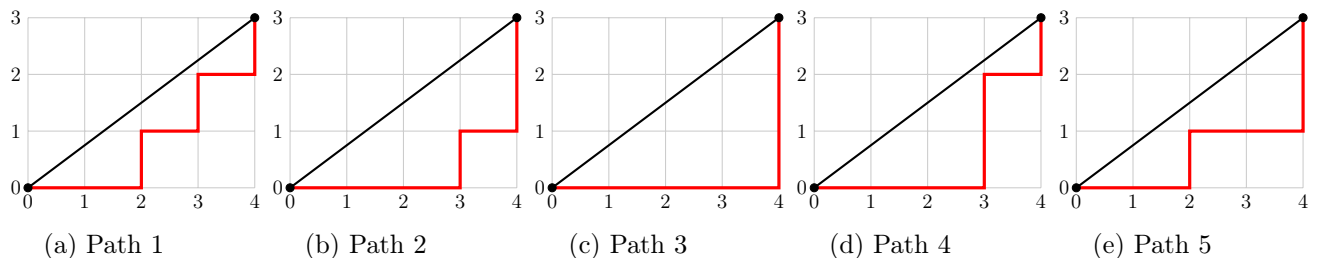
Exercise 7.1:

Is the counting of Dyck paths for given value of n related to any of the previous exercises?

Exercise 7.2:

Suppose there is a point (c, d) on the line segment joining $(0, 0)$ and (a, b) , where a, b, c, d are all positive natural numbers. What can you say about the relationship between a, b, c and d ?

Now let us look at all step paths not crossing the line segment from $(0,0)$ to (a, b) , where a, b are positive integers. The following figure shows all such paths for $(a, b) = (4, 3)$.



Assume a, b are relatively prime. A *Christoffel path* is the step path that hugs the line segment, i.e., there is no point with integer coordinates in the region between the line segment and the step path. Only Path 1 in the above figure is a Christoffel path. The corresponding word $xxxyxyxy$ is known as the *Christoffel word* of slope $\frac{3}{4} = \frac{\text{number of } y\text{s}}{\text{number of } x\text{s}}$.

Path 2 is not a Christoffel path since $(2, 1)$ lies in the area between the path and the line segment.

Exercise 7.3:

- a) Find Christoffel words of slopes $\frac{3}{4}$, $\frac{2}{5}$ and $\frac{3}{2}$.
- b) What can you say about first and last letter in a Christoffel word?
- c) Can you guess the relationship between Christoffel words of slopes $\frac{a}{b}$ and $\frac{b}{a}$?
- d) Show that there is a unique Christoffel word for a given (a, b) .