



---

## PoTW 2: Week of 6-3-2021 (solution)\*

Problem of the Week at [shsmathteam.com](https://shsmathteam.com)

---

### Problem of the Week #2: Catalan Crazyiness

*Topic: Combinatorics*

*Source: Andruwu Liuwu*

- (a). Based on the formula for the  $n$ th Catalan number, calculate the number of bad sequences of length  $2n$ , in terms of  $n$ .
- (b). Any sequence which starts with 1  $X$  and 2  $Y$ 's, in some order, must be bad. Calculate the number of sequences comprising of  $n$   $X$ 's and  $n$   $Y$ 's which start with 1  $X$  and 2  $Y$ 's (in some order), and then do the same for sequences comprising of  $n - 1$   $X$ 's and  $n + 1$   $Y$ 's. Show that these quantities are equal.
- (c). Finish the proof for the given formula for the  $n$ th Catalan number.
- (d). Prove that

$$C_n^{(2)} = \frac{1}{2n+1} \binom{3n}{n}.$$

- (e). Find and prove a formula for  $C_n^{(m)}$ , the  $m$ th-order Catalan number.

---

\*For inquiries: [andliu22@students.d125.org](mailto:andliu22@students.d125.org)

We present our solution in multiple parts:

**Solution, part (a):** The total number the sequences of length  $2n$ , such that  $n$  of them are  $X$ 's, and  $n$  of them are  $Y$ 's, is  $\binom{2n}{n}$ . Therefore, the number of bad sequences is equal to

$$\begin{aligned} \binom{2n}{n} - \frac{1}{n+1} \binom{2n}{n} &= \frac{n}{n+1} \binom{2n}{n} \\ &= \frac{n}{n+1} \cdot \frac{(2n)!}{n!n!} \\ &= \frac{(2n)!}{(n-1)!(n+1)!} \\ &= \binom{2n}{n-1}. \end{aligned}$$

As mentioned in the problem, we are motivated to represent the number of bad sequences in this way because it gives us a solid connection between the number of bad sequences, and the total number of sequences with  $n-1$   $X$ 's and  $n+1$   $Y$ 's. Therefore, if we can find a bijection between the two sets of sequences, then we would be finished with our proof. Part (b) allows us to explore some of the patterns which would ultimately lead to our bijection:

**Solution, part (b):** For the number of sequences consisting of  $n$   $X$ 's and  $n$   $Y$ 's, then the given starting pattern is such that the remainder of our sequence has  $n-1$   $X$ 's and  $n-2$   $Y$ 's. Because there are  $2n-3$  total letters to arrange after the starting 3, there are thus  $\binom{2n-3}{n-1}$  ways to arrange the remaining portion of our sequence. Similarly, for the number of sequences consisting of  $n-1$   $X$ 's and  $n+1$   $Y$ 's, then the given starting pattern is such that the remainder of our sequence has  $n-2$   $X$ 's and  $n-1$   $Y$ 's, so that there are also  $\binom{2n-3}{n-1}$  ways to arrange the remaining portion of our sequences. Finally, because the number of ways to arrange the starting three letters is the same in both instances, then we can thus conclude that the number of sequences with  $n$   $X$ 's and  $n$   $Y$ 's, and the number of sequences with  $n-1$   $X$ 's and  $n+1$   $Y$ 's, whose first three letters are some arrangement of 1  $X$  and 2  $Y$ 's, are equal.

Having established this specific relationship for any sequence starting with 1  $X$  and 2  $Y$ 's, we might now conjecture that this relationship holds for any bad sequence; that is, this should hold true for any sequence starting with  $k$   $X$ 's and  $k+1$   $Y$ 's, for  $0 \leq k \leq n-1$ . This allows us to complete our bijection and finish the proof:

**Solution, part (c):** We will show that the above proof actually holds regardless of the starting configuration. More specifically, suppose that we want to count the number of sequences with an arbitrary starting subsequence  $S_k$ , which consists of  $k$   $X$ 's and  $k+1$   $Y$ 's, for some  $0 \leq k \leq n-1$ .

For sequences which consists of  $n$   $X$ 's and  $n$   $Y$ 's total, whose first  $2k+1$  letters are the subsequence  $S_k$ , the remainder of the sequence (the part of the sequence which is not  $S_k$ ) must have  $n-k$   $X$ 's and  $n-k-1$   $Y$ 's, for a total of  $\binom{2n-2k-1}{n-k}$  total ways to arrange the rest of the sequence. In a similar way, for sequences which consists of  $n-1$   $X$ 's and  $n+1$   $Y$ 's total, whose first  $2k+1$  letters are the subsequence  $S_k$ , the remainder of the sequence must have  $n-k-1$   $X$ 's and  $n-k$   $Y$ 's, again for a total of  $\binom{2n-2k-1}{n-k}$  total ways to arrange the rest of the sequence.

Now, note that all bad sequences with  $n$   $X$ 's and  $n$   $Y$ 's must start with some configuration of  $\ell$   $X$ 's and  $\ell+1$   $Y$ 's, for  $0 \leq \ell \leq n-1$ . Therefore, in our above paragraph, by summing over all possible  $S_k$ , we have shown that the number of bad sequences with  $n$   $X$ 's and  $n$   $Y$ 's, is equal to the number of bad sequences

with  $n - 1$   $X$ 's and  $n + 1$   $Y$ 's. But, because *all* sequences with  $n - 1$   $X$ 's and  $n + 1$   $Y$ 's are bad, then the number of bad sequences with  $n$   $X$ 's and  $n$   $Y$ 's is just equal to the total number of sequences with  $n - 1$   $X$ 's and  $n + 1$   $Y$ 's, or  $\binom{2n}{n-1}$ . Therefore, we are done by part (a).

For the sake of brevity, we will not provide a solution to part (d), because part (d) is implied from the case when  $m = 2$  in our proof for part (e). With regards to our actual solution to the general case, we take major inspiration from the hint, and our proof will follow almost exactly the same as in parts (a), (b), and (c) (in fact, much of it will be copied verbatim).

**Solution, part (e):** We claim that the answer is

$$C_n^{(m)} = \frac{1}{mn+1} \binom{(m+1)n}{n}.$$

We will first count the number of  $m$ -bad sequences, which have  $n$   $X$ 's, and  $mn$   $Y$ 's. Let  $S_k$  be an arbitrary subsequence which consists of  $k$   $X$ 's and  $km + 1$   $Y$ 's, for some  $0 \leq k \leq n - 1$ .

For sequences which consists of  $n$   $X$ 's and  $mn$   $Y$ 's total, whose first  $(m+1)k+1$  letters are the subsequence  $S_k$ , the remainder of the sequence (the part of the sequence which is not  $S_k$ ) must have  $n - k$   $X$ 's and  $mn - (km + 1)$   $Y$ 's. Therefore, there is a total of

$$A_1^{(k)} = \binom{(m+1)n - (m+1)k - 1}{n - k}$$

total ways to arrange the rest of the sequence.

Now consider a similar line of reasoning, but this time for sequences which consist of  $n - 1$   $X$ 's and  $mn + 1$   $Y$ 's total. Suppose that we want to count the number of such sequences which also begin with the subsequence  $S_k$ . Then, the remainder of such a sequence (the portion of the sequence which is not  $S_k$ ) must have  $(n - 1) - k$   $X$ 's and  $(mn + 1) - (km + 1)$   $Y$ 's, which would therefore give us a total of

$$A_2^{(k)} = \binom{(m+1)n - (m+1)k - 1}{n - k - 1}$$

ways to arrange the rest of the sequence.

The key claim is that  $A_1^{(k)} = m \cdot A_2^{(k)}$  for all  $k$ , which we can show algebraically:

$$\begin{aligned} A_1^{(k)} &= \binom{mn - mk + n - k - 1}{n - k} \\ &= \frac{(mn - mk + n - k - 1)!}{(n - k)!(mn - mk - 1)!} \\ &= \frac{mn - mk}{n - k} \left[ \binom{n - k}{mn - mk} \left( \frac{(mn - mk + n - k - 1)!}{(n - k)!(mn - mk - 1)!} \right) \right] \\ &= m \left[ \frac{(mn - mk + n - k - 1)!}{(n - k - 1)!(mn - mk)!} \right] \\ &= m \binom{mn - mk + n - k - 1}{n - k - 1} \\ &= m \cdot A_2^{(k)}. \end{aligned}$$

Having established the above relationship, we are now ready to finish our proof. Note that all  $m$ -bad sequences with  $n$   $X$ 's and  $mn$   $Y$ 's must start with some configuration of  $\ell$   $X$ 's and  $m\ell + 1$   $Y$ 's, for

$0 \leq \ell \leq n-1$ . Therefore, by the relationship established above, by summing  $A_1^{(k)}$  and  $A_2^{(k)}$  over all possible  $S_k$ , we have shown that the number of  $m$ -bad sequences with  $n$   $X$ 's and  $mn$   $Y$ 's, is equal to  $m$  times the number of  $m$ -bad sequences with  $n-1$   $X$ 's and  $mn+1$   $Y$ 's. But, because *all* sequences with  $n-1$   $X$ 's and  $mn+1$   $Y$ 's are  $m$ -bad, then the number of  $m$ -bad sequences with  $n$   $X$ 's and  $mn$   $Y$ 's is precisely equal to  $m$  times the total number of sequences with  $n-1$   $X$ 's and  $mn+1$   $Y$ 's, which is  $m \binom{(m+1)n}{n-1}$ . Finally, we can now calculate the total number of  $m$ -good sequences with  $n$   $X$ 's and  $mn$   $Y$ 's, which is equal to the total number of such sequences (bad or good), minus the total number of such sequences which are  $m$ -bad:

$$\begin{aligned} C_n^{(m)} &= \binom{(m+1)n}{n} - m \cdot \binom{(m+1)n}{n-1} \\ &= \binom{(m+1)n}{n} - m \cdot \left[ \frac{n}{mn+1} \binom{(m+1)n}{n} \right] \\ &= \left( 1 - \frac{mn}{mn+1} \right) \binom{(m+1)n}{n} \\ &= \frac{1}{mn+1} \binom{(m+1)n}{n}, \end{aligned}$$

as desired.

**Final remarks:** With regards to authorship, the entire structure of the problem was inspired heavily by "Double-Good Catalan", an article written by Richard Rusczyk for WOOT.

This problem stood out to me because it (partially) resolves what feels like a really natural extension to the Catalan numbers. Not only are they often tied to the analogy of Dyck words (which is the formal name for the sequences described in this week's problem), but the  $n$ th Catalan number is also associated with myriad of other famous equivalent problems, including counting the number of lattice paths in a square region of length  $n$  which do not cross its diagonal (the line  $y = x$ ), and also counting the number of ways to triangulate a polygon of order  $n+2$ . But less often discussed are the obvious extensions to these analogies; for example, in the path analogy, what if the requirement shifts to paths which lie below the line  $y = 2x$ , instead of  $y = x$  (this corresponds to the number of 2-good sequences in part (d) of our problem); or, more generally,  $y = mx$  for any arbitrary slope  $m$ ? The introduction of these sorts of generalizations was what mainly motivated me to pick this week's problem, although it was admittedly quite long and more involved than I had originally intended.

For further exploration, the generalizations discussed by this week's problem can actually be taken into two dimensions, by increasing the number of  $X$ 's allowed per  $Y$ . For example, instead of 2  $Y$ 's for every  $X$ , as in part (d), we could have also explored what happens when we allow for 3  $Y$ 's for every 2  $X$ 's, which in the slope example would correspond with the number of paths not crossing the line  $y = 3/2x$ . This two-dimensional generalization gives way to the **Fuss-Catalan numbers**. In the notation of the linked Wikipedia article,  $r = 1$  was the specific case discussed by this week's problem.