Catalan Multijections

Garth Isaak[*], Larry Langley

Draft of September 16, 2018

Our aim here is to discuss variants on certain partition proofs that the Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}$ count various families of combinatorial objects. These are proofs that partition a size $\binom{2n}{n}$ set into parts of size $n+1$, each part containing exactly one element from some Catalan collection of objects. In a slight abuse of terminology we will call these multijective proofs. We will explore when they are 'different' from or a 'translation' of other multijective proofs. We hope this will encourage others to find other interesting multijective proofs of the Catalan count to parallel the plethora of bijective proofs between Catalan classes of objects.

One of the many examples enumerated by Catalan numbers are well formed sequences of parenthesis. These are sequences of n left (and n right) for which each initial segment has at least as many left parenthesis as right. Consider an arbitrary sequence of $n = 5$ open and 5 closed parentheses, but not a well formed one, say,)())((())(. We will follow a procedure that will identify a unique well formed sequence. We can identify well formed pairs by iteratively removing immediate open-closed pairs:

$$)\Big(\Big))((\Big(\Big))(\Rightarrow)(((\Big(\Big)($$

. After identifying those pairs, we then reverse all remaining) and remaining (.

$$\Big)()\Big)\Big((())\Big(\Rightarrow\Big(()\Big(\Big)(())\Big)$$

This procedure converts an arbitrary arrangement of parenthesis into a well formed arrangement, where each open parenthesis is later followed by its corresponding closed parenthesis.

We can reverse this procedure as well, by creating a collection of possible sequences that would result in the single well formed original. Consider the well formed arrangement of parentheses ()((()())). For each of the $n = 5$ paired () take the pair and all other pairs containing it and swap () to)(to get a new string of 5 left and 5 right parentheses. The

[*]Department of Mathematics, Lehigh University, Bethlehem, PA 18015 gisaak@lehigh.edu

5 new arrangements are shown below. We indicate from left to right on each line, using larger sizes to indicate the selected pair first; then the selected pair and pairs containing it; and finally the result after the swap:

$$\Bigg(\Bigg)((()())) \Rightarrow \Bigg(\Bigg)((()())) \Rightarrow \Bigg)\Bigg(((()()))$$

$$()\Bigg((()())\Bigg) \Rightarrow ()\Bigg((()())\Bigg) \Rightarrow ()\Bigg)(()())\Bigg($$

$$()(\Bigg(()()\Bigg)) \Rightarrow ()\Bigg(\Bigg(()()\Bigg)\Bigg) \Rightarrow ()\Bigg)\Bigg)()()\Bigg(\Bigg($$

$$()((\Bigg(\Bigg)())) \Rightarrow ()\Bigg(\Bigg(\Bigg(\Bigg)()\Bigg)\Bigg) \Rightarrow ()\Bigg)\Bigg)\Bigg)\Bigg(()\Bigg(\Bigg($$

$$()((()\Bigg(\Bigg))) \Rightarrow ()\Bigg(\Bigg(()\Bigg(\Bigg)\Bigg)\Bigg) \Rightarrow ()\Bigg)\Bigg)()\Bigg)\Bigg(\Bigg(\Bigg($$

For a given well formed arrangement of n right and left parentheses the n resulting arrangements are distinct. Furthermore, applying this to all well formed arrangement, all results are distinct. By including the original with its n results we partition the $\binom{2n}{n}$ arrangements of n left and right parentheses into parts of size $n+1$ each containing exactly one well formed arrangement. With a formal proof of this idea for the general case we would establish that the number of well formed arrangements of parentheses is the Catalan number $\frac{1}{n+1}\binom{2n}{n}$.

Rubenstein [15] describes this multijective proof on well formed arrangements of parentheses. Because of notational differences his construction appears different, making use of objects called inner and outer nests. However, a careful examination shows that the inverse multijection we have just described is indeed the same as Rubenstein's. Our approach seems to be a easier to describe. We also follow the approach of [15] in describing the other direction, the multijection taking arbitrary sequences to well formed sequences. We will look at this inverse more carefully later.

Our goal here is to focus on the potential for various simple multijective proofs for families of combinatorial objects counted by Catalan numbers. We will not discuss other standard approaches, using recursion and generating functions and using the reflection principle. Our focus is on one particular aspect of the counting problem. We note that there are a number of books or portions of books denoted to Catalan numbers. Four such are [7], [11], [12] and [17].

1 Definitions and History

1.1 Mulitjections

One of the first counting arguments students encounter in a combinatorics class determines the number $\binom{n}{k}$ of k element subsets of an n set by associating with each k subset its $k!$

permutations. These collections partition the $\frac{n!}{(n-k)!}$ length k permutations of an n set into parts of size $k!$ establishing that $\binom{n}{k} = \frac{1}{k!} \cdot \frac{n!}{(n-k)!}$. Our goal here is to examine proofs that parallel this structure in the Catalan setting.

For easy reference we will say that a map α from A to $\mathcal{A}$ is a t uniform multijection if the preimages $\alpha^{-1}(a)$ all have size t. Then $|\mathcal{A}| = \frac{1}{t}|A|$. In the subset example α maps length k permutations to the underlying set and α^{-1} associates a set with the $k!$ permutations of the set. We will drop t uniform from multijection as all of our examples will be $n+1$ uniform.

1.2 Catalan Families

There are hundreds of families of objects counted by the Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}$. The book [17] includes descriptions of bijections between many of these families. In most of our following examples will primarily make reference to two of these families. Ballot lists are strings of n 0's and n 1's where each initial substring has at least as many 1's as 0's. We will refer to the larger family of $\binom{2n}{n}$ binary strings with n 0's and n 1's as balanced binary strings. We will refer to Catalan multisets as size n multisets from $\{0, 1, \ldots, n\}$ dominated by $\{0, 1, 2, \ldots n\}$ when listing elements in nondecreasing order $a_1 \leq a_2 \leq a_n$. That is, we have $a_i \leq i - 1$ for all i. For notational simplicity we drop parentheses and commas and write the multset as a nondecreasing string. We will refer to the larger family of the $\binom{n+1+n-1}{n} = \binom{2n}{n}$ size n multisets from $\{0, 1, \ldots, n\}$ simply as multisets as the size and ground set will remain the same. We will refer to the standard bijection as the one that maps a binary string to a multiset by recording the the number of 0's preceding each 1.

Although well formed sequences of parenthesis do not seem to appear explicitly in [17] there is a trivial bijection to Ballot lists by identifying 1's with (and 0's with). However viewing these as parentheses provides a natural pairing of the parentheses that will be useful. In [13] these are called CG-arrangements after John Conway and Richard Guy. We will use the terms well formed CG-arrangements and will refer to the larger family of all strings of n left (and n right) as general CG-arrangements.

We note that bracketings which appear in [17] also involve n left and right parentheses are ways to use parenthesis to multiply $n+1$ numbers. These are different from CG-arrangements, which use parentheses but no other symbols. While both are counted by Catalan numbers, the underlying parentheses strings for well formed CG-arrangements and bracketings are different. See [13] for a bijection between these two families.

1.3 Mulijective Proofs

There are two standard mulitjective approaches to Catalan numbers, frequently called the cycle lemma of Dvosky and Motzkin and the Chung-Feller-MacMahon Theorem. See for example, [8], [14] or any of the books books mentioned previously. We will refer to these as CDM and CFM respectively. Each of these goes well beyond the basic counting problem. We will omit details of the additional structure as well as formal proofs as they can be found in many standard references. We will provide only brief descriptions and examples of these approaches in order to understand our notion of 'different' multijective proofs. Examples will follow in the next section.

We will refer to the multijection on well formed CG-arrangements and general CG-arrangements of parenthesis described in the introduction as CGP.

The Dvorsky-Motzkin Cycle Lemma and the Chung-Feller-Macmahon Theorem are typically phrased in terms of equivalence classes or uniform partitions each part of which contains a Catalan object and in the case of CFM there are additional properties that provide a further refinement and a second 'orthogonal' partition. In these cases the collection of Catalan objects is a subset of a larger set of size $\binom{2n}{n}$. For example, when described in terms of Ballot lists or well formed CG-sequences we have the inclusions noted above.

We prefer the multijection terminology since it fits better with the possibility of proofs where the Catalan object has no such natural embedding. For example, binary trees are Catalan objects but there is no obvious larger class of $\binom{2n}{n}$ trees for which the binary trees are a special case. However, multijective proofs using binary trees on n vertices to associated binary strings with n 0's and n 1's might be interesting. See [2], [4] for examples multijection examples related to plane trees.

For our purposes CDM can be briefly described as follows in terms α^{-1} for a multijection. Given a ballot list, append an additional 1 at the beginning, write down all cyclic shifts starting with a 1 and then record these strings omitting the initial 1.

For our purposes CFM can be briefly described as follows. Given a mulitset written in nondecreasing order consider the lists obtained by adding k to each term, reducing modulo $n+1$ and rearranging to get a multiset listed in nondecreasing order. This will suffice for our counting purposes even though it does not capture the full generality of CFM. The version we have given is found for example in [12], however the are different proofs of the full CFM results as will be noted in Section 4.4.

2 Examples

For illustration we provide examples of three multijections, CDM, CFM and CGP for the case $n = 3$. This will help motivate some of our later ideas of different and translation.

There are $C_3 = \frac{1}{3+1}\binom{6}{3} = 5$ ballot lists: $111000, 110100, 110010, 101100, 101010$; 5 corresponding Catalan multisets $000, 001, 002, 011, 012$; and 5 corresponding well formed CG-arrangements $((())), (()()), (())(), ()(()), ()()()$.

2.1 CDM

To illustrate CDM as as described above we find the inverse image of 110010 by first appending a 1 at the front and writing down all cyclic shifts of this string that start with a 1

$$\begin{array}{cccccccccccc}
1 & 1 & 1 & 0 & 0 & 1 & 0 & & & & & \\
 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & & & & \\
 & & 1 & 0 & 0 & 1 & 0 & 1 & 1 & & & \\
 & & & & & 1 & 0 & 1 & 1 & 1 & 0 & 0
\end{array}$$

Removing the lead 1 from these we get the inverse image of 110010 to be
$110010, 100101, 001011, 011100$.

In a similar manner we associate a set of 4 binary strings with three 1's and 3 0's with each ballot list. This partitions all $\binom{6}{3} = 20$ such strings into 5 preimages each of size 4 each of which contains exactly one ballot list. We write down the ballot lists at the top of each preimage. For comparison we also write down the corresponding multisets under the standard bijection.

111000	110100	110010	101100	101010
110001	101001	100101	011001	010101
100011	010011	001011	100110	010110
000111	001110	011100	001101	011010

000	001	002	011	012
003	013	023	113	123
033	133	233	022	122
333	222	111	223	112

2.2 CFM

To illustrate CFM as described above to find the inverse image of 002 we first add $0, 1, 2, 3$ modulo 4 to get $002, 113, 220, 331$ and then rearrange in nondecreasing order to write these multisets as $002, 113, 022, 133$.

In a similar manner we associate a set of 4 size 3 multisets from $\{0, 1, 2, 3\}$ with each Catalan multiset. This partitions all $\binom{4+3-1}{3} = \binom{6}{3} = 20$ such multisets into 5 preimages each of size 4 each of which contains exactly one Catalan multiset. We write down the Catalan multisets at the top of each preimage. For comparison we also write down the corresponding balanced binary strings under the standard bijection

000	001	002	011	012
111	112	113	122	123
222	223	022	233	023
333	033	133	003	013

111000	110100	110010	101100	101010
011100	011010	011001	010110	010101
001110	001101	100110	001101	100101
000111	100011	010011	110001	101001

2.3 CGP

To illustrate CGP as described above we find the inverse image of $(())()$ in a manner similar to our introductory example:

$$\Big(()\Big)() \Rightarrow \Big(()\Big)() \Rightarrow \Big)()\Big(()$$

$$(\Big(\Big))() \Rightarrow \Big(\Big(\Big)\Big)() \Rightarrow \Big)\Big)\Big(\Big(()$$

$$(())\Big(\Big) \Rightarrow (())\Big(\Big) \Rightarrow (())\Big)\Big($$

Along with the intial set (corresponding to picking no parenthesis) we get the following preimage: $(())(),)()((),))((((), (()))($.

In a similar manner we associate a general CG-arrangement with each well formed CG-arrangement. This partitions all $\binom{6}{3} = 20$ CG-arrangements into 5 preimages each of size 4 each of which contains exactly one well formed CG-arrangement. We write down the well formed CG-arrangements at the top of each preimage.

((()))	(()())	(())()	()(())	()()()
)(())(	)()()(	)()(()	)((())	)(()()
))()((	))(()(	))((()	())()(	())(()
)))(((	)())((	(()))(	()))((	()())(

For comparison we write down the corresponding balanced binary strings and and mutlsets using the standard bijections:

111000	110100	110010	101100	101010
011001	010101	010110	011100	011010
001011	001101	001110	100101	100110
000111	010011	110001	100011	101001

000	001	002	011	012
113	123	122	111	112
233	223	222	023	022
333	133	003	033	013

2.4 Comparing the Partitions

Referring back to the three multijective examples it appears that they produce different partitions, at least when $n = 3$. However, if we examine the binary strings we observe that taking the complement of each string (swapping 0's and 1's) we do get the same partition for CFM and CDM. In particular the complement of the first and fifth groups under CDM are the first and fifth groups respectively under CFM. The complement of the second CDM group is the fourth CFM group, the complement of the third CDM group is the second CFM group and the complement of the fourth CDM group is the third CFM group.

We are not aware of any simple connection between the CG partitions and either CDM or CFM although it would be interesting to describe one.

3 Different vs Translation

In order to discuss whether CDM and CFM as presented in the examples above are 'different' in some sense or more like 'translations' into different notation we use the following.

Let α be a multijection from A to $\mathcal{A}$ and β a multijection from B to $\mathcal{B}$ and let σ be a bijection from A to B and π be a bijection from $\mathcal{A}$ to $\mathcal{B}$. Hence $|A| = |B|$ and $|\mathcal{A}| = |\mathcal{B}|$ and the sizes of the preimages are the same under both multijections. We will say that α and β are translations of each other under σ and π if the maps under the bijections are the same. That is, $\pi(\alpha(a)) = \beta(\sigma(a))$ for all $a \in A$. In particular, under σ the preimages induce the same partition.

....Put a diagram figure here ... comment that it is a commutative diagrams??

This still does not provide a technical definition of 'different' or 'translation' without the context of the bijections. However, we would informally tend to think of two multijections as translations of each other if they are translations under some 'natural' or 'standard' choice of σ or π. For example if $A = B$ and σ is the identity and similarly for π. In the examples above, using the standard bijection between n multisets and balanced binary strings CDM and CFM are different but if σ is complementation of the binary strings followed with the standard bijection then CDM and CFM are translations of each other. In the next section we describe the general connection for any n that puts this particular translation in perspective.

In our examples CGP appears to be different under obvious choices for the bijections. However perhaps closer comparison will produce an interesting bijection between balanced binary strings or multisets and general CG-arrangements under which these multijections are translations.

4 Proof Sketches

4.1 CGP multijection

In the introduction we described α^{-1} for a mulitjection from CG-arrangements to well formed CG-arrangements. The inverse image of well formed CG-arrangement with n pairs of parentheses consists of the arrangement itself along with n arrangements formed by taking each of the n pairs and swapping (and) in that pair along with all pairs that contain it.

We also provided an example of α following the approach of [15]. Given a general CG-arrangement, iteratively remove adjacent pairs of a left (immediately followed by a right). What remains is a sequence of k right) followed by k left (for some k. For these remaining parentheses, change all (to) and change all) to (and insert into the original arrangement. It is straightforward to check that all of the removed pairs fall together in some gap of the changed pairs, hence the process of removing adjacent () on the new arrangement removes all pairs. The new arrangement is well formed with the original resulting from the choice of the innermost pair of the k that were changed (or from itself if the original is well formed and $k = 0$).

4.2 Connection between CFM and CDM

As observed in the examples it seems that CFM and CDM are translations when σ maps balanced binary strings to multisets by first complementing the string, swapping 0's with 1's and then following the standard bijection.

To outline a proof that this is indeed the case in general we first recall the stars and bars representation for multisets. We associate a k element multiset $a_1 \leq a_2 \leq \cdots \leq a_k$ of $\{0, 1, \ldots, n\}$ with a string of n bars, $|$'s and k stars, $*$ by placing $*$'s in positions $a_i + i$ for $i = 1, 2, \ldots, k$. The n bars form $n + 1$ bins with the number of stars in each bin the count of the corresponding element in the multiset.

If we replace the 0's in a binary string with stars and think of the 1's as bars the corresponding multiset is that obtained by first complementing and then applying what we called the standard bijection. Equivalently, we count the number of 1's preceding each 0. This is also a natural bijection which might have an equal claim to be the 'standard' bijection. Note however that this bijection does not map Ballot lists to Catalan multisets while our standard bijection does. So for our translation π will remain the standard bijection while σ will be as described in this paragraph.

In our case we have size n multisets from $\{0, 1, \ldots, n\}$ and hence strings of n stars and n bars. The preimage family for a balanced binary string is obtained by appending a lead 1, taking all cyclic shifts starting with a 1 and for each deleting the lead 1. This was the construction applied to a Ballot list in the example of Section 2.1. The cycle lemma tells us that exactly one string in a preimage family will be a Ballot list which provides a path to describing the map α directly. See, for example, [6] for nice descriptions of this.

If we examine the cyclic shifts from the example in Section 2.1 and apply σ as described in the previous paragraph, we see that the number of 1's preceding a given 0 decreases by 1 modulo $n+1$ for each shift. So counting the 1's preceeding each 0, the strings $110010, 100101, 001011, 011100$ obtained from the shifts under σ are $223, 112, 001, 330$ with the last written as in order as 033. This corresponds exactly to the middle preimage in the CFM example of Section 2.2.

It is not hard to see that this works in general, the cyclic shifts in the CDM families correspond exactly to subtracting 1 from each count of 1's preceding a 0, i.e. subtracting 1 modulo $n+1$ from each element in the multiset. So from this perspective, with this σ and π we view CDM and CFM as described in our examples as translations of each other.

4.3 Translations and Mulitjections

We can also use the description of translations to rephrase multijections in other terminology. As one example consider CDM keep σ as the standard bijection, rather than complementation followed by the standard bijection as above. We can then ask what β on multisets would be under this translation. This allows us to discover a different mulitjective proof on multisets.

We describe this briefly by example, without proof.

Given a Catalan multiset from $\{0, 1, \ldots, n\}$, $a_1 \leq a_2 \leq \cdots \leq a_n$ create a new string $b_1 \leq b_2 \leq \cdots \leq b_n$ by setting $b_i = a_{i+1} - a_1$ for $i = 1, 2, \ldots, n-1$ and $b_n = n - a_1$. Repeat this n times to get the preimage (including also the original multiset). Observe that this same process works for any n multiset so applying it to any n multiset we get a set of $n+1$ multisets containing exactly one Catalan multiset.

As an alternative description, the multiset obtained by applying the above j times has $c_i = a_{i+1} - a_j$ for $i = 1, 2, \ldots, n-j$ and $c_{n-j+1} = n - a_j$ and $c_i = n - a_j + a_x$ for $i = n-j+2, \ldots, n$ with $i = n-j+1+x$

For example consider 002 when $n = 3$. The first additional multiset in the inverse image is $0-0, 2-0, 3-0 = 023$, the second is $2-0, 3-0, 3-0 = 233$, the third is $3-2, 3-2, 3-2 = 111$. If we were to iterate one more time we get $1-1, 1-1, 3-1 = 002$, the starting Catalan multiset. Note that this is indeed the family we obtained from 110010 using CDM in Section 2.1.

4.4 Other CDM and CFM proofs

We note that the cycle lemma is also sometimes viewed with the appended 1 as part of the family of objects. That is, we consider strings with $n+1$ ones and n zeros and view the Catalan numbers as $C_n = \frac{1}{2n+1}\binom{2n+1}{n}$. We leave as an exercise the translation, similar to that above, into size $n = 1$ multisets from $\{0, 1, \ldots, n\}$, where the preimage families now have size $2n+1$. A nice geometric interpretation of the cycle lemma is in [6]. Some examples of CDM proofs and generalizations include [3], [9], [16].

Proofs of Chung-Feller-Macmahon are frequently given in terms of other objects, in particular Dyck paths. The broader context creates a different partition of a family of size $\binom{2n}{n}$, into $n+1$ parts each of size C_n, with exactly one of these parts consisting of the Catalan family. Each part has a common value of some statistic, for the example of balanced binary string, the largest count for numbers of 0's preceding a 1. The Ballot lists are the part where this value is 0. What we have called preimages are an 'orthogonal' partition in the sense that each preimage contains exactly one object for a given value of the statistic. See [1] for a description of such a statistic on the CGP families. See, for example, [5], [10], [18] for examples of CFM proofs and generalizations.

For our purposes we note that different proofs of CFM in fact describe different preimage partitions. For example, it is not difficult to check with small examples that under natural bijections σ and π the CFM preimage families described in [10], [14], [18] are different from the one we have described. Indeed [10] explicitly creates a different preimage partition so that strings in a part differ only in two locations. It might be interesting to determine if there is a somewhat natural σ and π under which these different proofs are translations of each other. It might also be interesting to describe the translation under natural bijection to other objects, as we did in Section 4.3 for multisets and the version of CFM used here above.

5 Conclusion

Our goal here was to examine different versions of what we call multijective proofs for Catalan numbers. The framework sets up a large number of possible exercises: describing how known mulitjective proofs translate to other families of Catalan objects, describing bijections that make two seeming different multijective proofs into translations, and finding new multijective proofs that are different under natural choices for the bijections.

References

[1] Callan, D., Two Uniformly Distributed Parameters Defining Catalan Numbers, Amer. Math. Monthly **106** (1999), 948–949.

[2] Chen, W.Y.C, Nelson, Y.L., and Shapiro, L.W., The Butterfly Decomposition of Plane Trees, Discrete Appl. Math., **155** (2007), 2187–2201.

[3] Dershowitz, N., and Zaks, S., The Cycle Lemma and Some Applications, European J. Combin., **1** (1990), 35–40.

[4] Feldman, D., Counting Plane Trees, unpublished manuscript, 1992. Linked under A006082 and A006936 in On-Line Encyclopedia of Integer Sequences.

[5] Huq, A., Generalized Chung-Feller Theorems for Lattice Paths, Ph.D. Dissertation, Brandeis Univ. 2009.

[6] Graham, R., Knuth, D. and Patashnik, O., Concrete Mathematics, Addison-Wesley, 1989.

[7] Koshy, T., Catalan Numbers with Applications, Oxford Univ. Press, 2009.

[8] Krattenthaler, C., Lattice Path Enumeration, Chapter 10 in Handbook of Enumerative Combinatorics, CRC press, 2015. 589–678.

[9] Ma, J., Shen, H., and Yeh, Y., Rooted Cyclic Permutations of Lattice Paths and Uniform partitions, Disc. Math., **338** (2015), 1111-1125.

[10] Mütze, T., Standke, C. and Wiechert, V., A Minimum-Change Version of the Chung-Feller Theorem for Dyck Paths, European J. Combin., **69** (2018), 260–275.

[11] Mohanty, S.G., Lattice Path Counting and Applications, Academic press, 1979.

[12] Narayana, Latice Path Combinatorics, with Statistical Applications, Univ. Toronto press, 1979.

[13] Penrice, S., Stacks, Bracketings, and CG-Arrangements, Math. Mag., **72** (1999), 321–324.

[14] Renault, M., Four proofs of the Ballot Theorem, Math. Mag., **80** (2007), 345–352.

[15] Rubenstein, D., Catalan Numbers revisited, J. Combin. Theory Ser. A, **68** (1994), 486–490.

[16] Snevily, H.S. and West, D.B., The Bricklayer Problem and the Strong Cycle Lemma, Amer. Math. Monthly **105** (1998), 131–143.

[17] Stanley, R.C., Catalan Numbers, Cambridge Univ. Press, 2015.

[18] Woan, W., Uniform Partitions of Lattice Paths and Chung-Feller Generalizations, Amer. Math. Monthly **108** (2001), 556–559.