GENERALIZED CATALAN NUMBERS FROM HYPERGRAPHS

PAUL E. GUNNELL

Abstract. The Catalan numbers \((C_n)_{n \geq 0} = 1, 1, 2, 5, 14, 42, \ldots\) form one of the most venerable sequences in combinatorics. They have many combinatorial interpretations, from counting bracketings of products in non-associative algebra to counting rooted plane trees and noncrossing set partitions. They also arise in the GUE matrix model as the leading coefficient of certain polynomials, a connection closely related to the plane trees and noncrossing set partitions interpretations. In this paper we define a generalization of the Catalan numbers. In fact we actually define an infinite collection of generalizations \(C_{n}^{(m)}\), \(m \geq 1\), with \(C_{n}^{(1)}\) equal to the usual Catalans \(C_n\); the sequence \(C_{n}^{(m)}\) comes from studying certain matrix models attached to hypergraphs. We also give some combinatorial interpretations of these numbers.

1. Introduction

1.1. The Catalan numbers \((C_n)_{n \geq 0}\)

\[1, 1, 2, 5, 14, 42, 132, 429, 1430, \ldots,\]

form one of the most venerable sequences in combinatorics. They have many combinatorial interpretations, far more than can be reproduced here. We only mention a few of the most famous ones:

(i) A plane tree is a rooted tree with an ordering specified for the children of each vertex. Then \(C_n\) counts the number of plane trees with \(n + 1\) vertices.

(ii) A Dyck path of length \(n\) is a directed path from \((0, 0)\) to \((n, 0)\) in \(\mathbb{R}^2\) that only uses steps of type \((1, 1)\) and \((1, -1)\) and never crosses below the \(x\)-axis. Then \(C_n\) counts the number of Dyck paths of length \(2n\).

(iii) A ballot sequence of length \(2n\) is a sequence \((a_1, \ldots, a_{2n})\) with \(a_i \in \{\pm 1\}\) with total sum 0 and with all partial sums nonnegative. Then \(C_n\) counts the number of ballot sequences of length \(2n\).

(iv) A binary plane tree is a plane tree in which every node \(N\) has at most two children, which are called the left and right children of \(N\). Furthermore, if

---

Date: 2 April 2019.

2010 Mathematics Subject Classification. Primary 05A10, 11B65.

Key words and phrases. Matrix models, hypergraphs, Catalan numbers.

The author was partially supported by NSF grant DMS 1501832. We thank Dan Yasaki and the referee for helpful comments.
a node has only one child, then it must be either a left or right child. Then $C_n$ counts the number of binary plane trees with $n$ vertices.

(v) $C_n$ is the number of ways to divide a regular $(n+2)$-gon into triangles without adding new vertices and by drawing $n-1$ new diagonals.

(vi) Let $\Pi$ be a polygon with $2n$ sides. Then $C_n$ counts the ways to glue the sides of $\Pi$ together in pairs to make an orientable topological surface of genus 0.

The definitive reference for combinatorial interpretations of Catalan numbers is Richard Stanley’s recent monograph [15]. It contains no fewer than 214 different interpretations of the $C_n$. Indeed, the first five interpretations given above appear in [15, Chapter 2] as items (6), (25), (77), (4), and (1) respectively. The last interpretation (counting genus 0 polygon gluings) is unfortunately not in [15]. However, it is easily seen to be equivalent to [15, Interpretation (59)], which counts the number of ways to draw $n$ nonintersecting chords joining $2n$ points on the circumference of a circle. Another resource is OEIS [12], where the Catalan numbers are sequence A000108.

1.2. The goal of this paper is to give a family of generalizations of the $C_n$. For each integer $m \geq 1$, we define a sequence of integers $(C_n^{(m)})_{n \geq 0}$; for $m = 1$ we have $C_n^{(1)} = C_n$. Here are some further examples:

$m = 2: 1, 1, 6, 57, 678, 9270, 139968, 2285073, 39871926, 739129374, 14521778820, \ldots$

$m = 3: 1, 1, 20, 860, 57200, 5344800, 682612800, 118180104000, 27396820448000, \ldots$

$m = 4: 1, 1, 70, 15225, 7043750, 6327749750, 10411817136000, 2903403169460625, \ldots$

$m = 5: 1, 1, 252, 299880, 1112865264, 11126161436292, 255654847841227632, \ldots$

$m = 6: 1, 1, 924, 6358044, 203356067376, 23345633108619360, \ldots$

$m = 7: 1, 1, 3432, 141858288, 40309820014464, 53321581727982247680, 238681094467043912358445056, \ldots$

The $C_n^{(m)}$ are defined in terms of sums over trees, weighted by the orders of their automorphism groups. For $m = 1$ the resulting expression is not usually given as a standard combinatorial interpretation of the Catalan numbers, but it is known to compute them; we will prove it in the course of proving Theorem 4.9. In fact, from our definition it is not clear that the $C_n^{(m)}$ are actually integers, even for $m = 1$, although we will see this by giving several combinatorial interpretations of them.

Here is the plan of the paper. In §2 we give the definition of the $C_n^{(m)}$, and in §3 we explain how to compute them for reasonable values of $n$ and any $m$. In §4 we give six different combinatorial interpretations of the $C_n^{(m)}$ based on six standard

\footnote{An earlier version of this list is contained in [14], with additions available on Stanley’s website [13].}
interpretations of the Catalan numbers. In §5 we explain how to compute the generating function of the $C^{(m)}_n$, and conjecture some asymptotics of $C^{(m)}_n$ for fixed $m$ as $n \to \infty$. Finally, in §6 we discuss how these numbers arise in the study of certain matrix models associated to hypergraphs.

2. Hypergraph Catalan Numbers

2.1. We begin by giving the description of the Catalan numbers that we wish to generalize. Let $\mathcal{T}_n$ be the set of unlabeled trees on $n$ vertices. The sequence $|\mathcal{T}_n|$ appears on OEIS as A000055, and begins

$$1, 1, 1, 2, 3, 6, 11, 23, 47, 106, \ldots,$$

where $|\mathcal{T}_0| := 1$ by convention.

Let $T \in \mathcal{T}_{n+1}$, and for each vertex $v \in T$, let $a_T(v)$ be the number of walks that (i) begin and end at $v$, (ii) visit every other vertex at least once, and (iii) traverse each edge of $T$ exactly twice, once while going away from $v$ and once while coming back to $v$. Let $\Gamma(T)$ be the automorphism group of $T$, and let $|\Gamma(T)|$ be its order.

Figure 1 shows an example of these numbers for the 3 trees in $\mathcal{T}_5$. For example, the outer numbers on the leaves of the upper left tree are 1 because the only possible walk is to go from one end of the tree to the other, then back to the beginning leaf. The inner numbers on the same tree are 2, because one must first choose a direction in which to head, then must go all the way to that end, then back through the starting point to the other end, then back to the initial vertex.

2.2. The connection between the trees and the Catalan numbers is given by the following result:

2.3. Proposition. The Catalan number $C_n$ is given by

$$C_n = \sum_{T \in \mathcal{T}_{n+1}} \sum_{v \in T} \frac{a_T(v)}{|\Gamma(T)|}.$$  

For example, using Figure 1 we have

$$\frac{8}{2} + \frac{48}{24} + \frac{16}{2} = 4 + 2 + 8 = 14 = C_4.$$  

We defer the verification of Proposition 2.3 until §4 when we discuss combinatorial interpretations. The proposition will be proved in the course of Theorem 4.9.

2.4. Now let $m \geq 1$ be a positive integer. Then $C^{(m)}_n$ is defined essentially as in (1), but we modify the definition of the numbers $a_T(v)$. For $T \in \mathcal{T}_{n+1}$, we let $a^{(m)}_T(v)$ count the walks that (i) begin and end at $v$, (ii) visit every other vertex at least once, and (iii) traverse each edge of $T$ exactly $2m$ times, $m$ times while going away from $v$ and $m$ times while coming back to $v$. 

Figure 1. The set $\mathcal{T}_5$ with vertices labeled by $a_T(v)$. Going clockwise from the upper left, the orders of the automorphism groups are 2, 24, 2.

2.5. Definition. The hypergraph Catalan numbers $C_n^{(m)}$ are defined by

\begin{equation}
C_n^{(m)} = \sum_{T \in \mathcal{T}_{n+1}} \sum_{v \in T} \frac{a_T^{(m)}(v)}{|\Gamma(T)|}.
\end{equation}

For example the numbers $a_T^{(2)}(v)$ are shown in Figure 2 for the three trees in $\mathcal{T}_5$. The numbers are larger now, since walks have many more options. Using the numbers in Figure 2 we find

\[ C_4^{(2)} = \frac{216}{2} + \frac{5040}{24} + \frac{720}{2} = 108 + 210 + 360 = 678. \]

Figure 2. The set $\mathcal{T}_5$ with vertices labeled by $a_T^{(2)}(v)$.

3. Computing the $C_n^{(m)}$

3.1. In this section we show that once one has sufficient knowledge of the trees in $\mathcal{T}_{n+1}$ to compute $C_n$, one can easily compute $C_n^{(m)}$ for any $m$. In other words, if one wants to extend the data in $\mathcal{T}_n$ to larger values, we show that it is easy to fix $n$ and let $m$ grow.
3.2. Theorem. Let $T \in \mathcal{T}_{n+1}$, and let $v \in T$ have degree $d(v)$. Then we have

$$
\sum_{v \in T} a^{(m)}_T(v) = \frac{2nm^{n+1}}{(m!)^2} \prod_{v \in T} (md(v) - 1)!.
$$

For example, for the tree on the right of Figure 2, we find

$$
2 \cdot 4 \cdot 2^5 \frac{(2!)^8 (2 \cdot 1 - 1)!4(2 \cdot 4 - 1)!}{(2!)^8} = 5040,
$$

which agrees with the data in the figure.

Proof. We use the results in [14, §5.6], which treats Eulerian tours in balanced digraphs, and so we begin by recalling some notation. Let $G$ be a connected digraph and let $\tilde{G}$ be the associated undirected graph. Suppose $G$ is balanced; this means that the outdegree $o(v)$ of each vertex $v$ is equal to its indegree $i(v)$. Given a vertex $v$ of $G$, an oriented spanning tree rooted at $v$ is a subgraph $T \subset G$ such that $\tilde{T}$ is a spanning tree of $\tilde{G}$ in the usual sense, and such that all the edges of $T$ are oriented towards $v$. Then since $G$ is connected and balanced it is Eulerian ([14, Theorem 5.6.1]), and by [14, Theorem 5.6.2] given an edge $e$ the number $\varepsilon(G, e)$ of Eulerian tours of $G$ beginning at $e$ is

$$
\varepsilon(G, e) = \tau(G, e) \prod_{v \in G} (o(v) - 1)!,
$$

where $\tau(G, e)$ is the number of oriented spanning trees of $G$ with root at the initial vertex of $e$. Furthermore, it is known that $\tau(G, e)$ is independent of $e$.

Now let $T$ be a tree with $n+1$ vertices and fix $m$. Let us say that an $a^{(m)}_T$-tour of $T$ is a closed walk on $T$ that traverses each edge exactly $2m$ times. We will count $a^{(m)}_T$-tours by first counting Eulerian tours on the canonical balanced digraph $T_m$ built from $T$ by replacing each edge with $2m$ edges, $m$ oriented in one direction and $m$ oriented in the other. Let $v \in T$. Then it is clear that a path contributing to $a^{(m)}_T(v)$ determines (non-uniquely) an Eulerian tour on $T_m$ starting and ending at $v$. Hence we can use (4) to compute $a^{(m)}_T(v)$. In particular for an edge $e \in T_m$ we have

$$
\varepsilon(T_m, e) = \tau(T_m, e) \prod_{v \in T} (md(v) - 1)!,
$$

where $d(v)$ is the degree of $v$ in $T$.

Now we go from (4) to (3). First, to get the total number of tours, we multiply by the number of edges, which is $2mn$. Next, the number $\tau(T_m, e)$ of oriented spanning trees in $T_m$ is $m^n$ (after fixing a root we pick one of $m$ possible properly oriented edges for each edge in $T$). Let $\pi$ be the map

$$
\pi: \{\text{Eulerian tours of } T_m\} \rightarrow \{a^{(m)}_T\text{-tours of } T\}
$$
that replaces each edge of an Eulerian tour of $T_m$ with the corresponding edge in $T$. It is clear that $\pi$ is surjective. Furthermore, each $a_T^{(m)}$-tour $w$ of $T$ has precisely $(m!)^{2n}$ preimages under this map, since $w$ traverses each edge of $T$ in each direction precisely $m$ times, and we get to choose in which order these $m$ traversals correspond to the $m$ corresponding edges of $T_m$. Thus

$$\left| \{ \text{Eulerian tours of } T_m \} \right| = (m!)^{2n} \cdot \left| \{ a_T^{(m)} \text{-tours of } T \} \right|.$$ 

This accounts for the $(m!)^{2n}$ in the denominator of (3), and completes the proof. □

3.3. Thus to compute $C_n^{(m)}$ for any $m$ one only needs the trees in $T_{n+1}$ together with their vertex degrees and orders of their automorphism groups. This can be done, at least for reasonable values of $n$, using the software nauty [10]. For example there are 751065460 $\approx 2^{29}$ trees in $T_{27}$; nauty is able to compute them on a laptop in less than 14 seconds. On the other hand, computing the orders of all the automorphism groups takes longer. For instance there are only 823065 $\approx 2^{20}$ trees in $T_{20}$, and computing all their automorphism groups takes just over 4 hours.

4. Combinatorial interpretations

4.1. As it turns out, the numbers $C_n^{(m)}$ have a variety of combinatorial interpretations, in fact in terms of objects used to count the usual Catalan number $C_n$. As we shall see, only certain of these will contribute to $C_n^{(m)}$, and in general a given object may contribute in several different ways. It will also be evident that any standard Catalan interpretation can be turned into one for the $C_n^{(m)}$. We begin by introducing some notation.

4.2. Let $X$ be an combinatorial object; we do not give a precise definition of $X$, but the reader should imagine that $X$ is something used in a standard Catalan interpretation, such as those in §1.1. We will give examples in Theorem 4.9. Typically $X$ will be an aggregate of smaller elements, and we say that a level structure for $X$ is a surjective map $\ell$ from these elements to a finite set $\{1, 2, \ldots, N\}$ for some $N \in \mathbb{Z}_{>0}$. We will say $x \in X$ is at a higher level than $x' \in X$ if $\ell(x) > \ell(x')$, with similar conventions for same and lower level. The $i$th level $X_i \subset X$ with respect to $\ell$ will be the preimage $\ell^{-1}(i) \subset X$. In some interpretations, we will also have a zeroth level $X_0$; these will usually be combinatorial objects that are naturally rooted. If $X$ has a zeroth level, we will require $|X_0| = 1$. Note that in the pictures that follow, the function $\ell$ will not typically correspond to the height of elements of $X$ in their positions in the figures.
4.3. The object $X$ will be a poset. If $x, x' \in X$ and $x$ covers $x'$, we will say that $x$ is a *parent* of the *child* $x'$. The level structures we consider will always be compatible with the poset structure, in that any child $x'$ of a given parent $x$ will satisfy $\ell(x') = \ell(x) + 1$. In other words, children will always lie on a higher level than their parents.

4.4. Finally, let $m$ be a positive integer. We will consider *$m$-labeling* the positive levels of $X$, which means the following. First fix a finite set $L$ of labels. Let

$$X = X_0 \sqcup \bigcup_{i \geq 1} X_i$$

be the disjoint decomposition of $X$ into levels, where $X_0$ may be empty. For each *positive* level $X_i$, $i > 0$, we choose a disjoint decomposition of $X_i$ into subsets of order $m$; in particular, this implies $|X_i| = 0 \mod m$ for $i > 0$ in all cases we consider. Then an $m$-labelling is a map from these subsets to $L$. We will say that an $m$-labeling is *admissible* if the following are true:

- Each subset receives a unique label.
- The labeling is compatible with the level structure, in the following sense: if two elements $x, x'$ share a given label, then the labels of their parents agree.

We also consider two labelings to be equivalent if one is obtained from the other by permuting labels.

4.5. We give an example to clarify this terminology. Let $X$ be a rooted plane tree. The elements of $X$ are its vertices. Let $x_0$ be the root. We can define a level structure $\ell: X \to \mathbb{Z}_{\geq 0}$ by setting $\ell(x)$ to be the distance in $X$ to $x_0$. Note that $X_0 = \{x_0\}$ has order 1, but of course the positive levels $X_i$ can be bigger. Given two vertices $x, x'$, one is a parent or a child of the other if it is in the usual sense of trees: $x$ is a parent (respectively, child) of $x'$ if $x$ and $x'$ are joined by an edge and $x$ lies closer to (respectively, further from) the root than $x'$. Figure 3 shows two rooted trees $X, Y$. We have embedded them in the plane with their roots the top of the picture. Each tree has four levels. Parents appear above their children, and levels increase as we move down the figure.

Next we consider labelings. Figure 4 shows the two rooted trees $X, Y$ equipped with 2-labelings, with labeling set $L = \{a, b, c, d, e\}$. We have arbitrarily ordered the labels as indicated; one would obtain an equivalent labeling after permuting the labels. The left tree $X$ is admissibly labeled: if two vertices have the same label, so do their parents. The right tree $Y$, however, is not admissibly labeled: the two vertices at the bottom have the same label $e$, but their parents have two different labels $c, d$. 
4.6. Now let $\mathcal{X}_{nm}$ be a set of objects constituting a combinatorial interpretation of $C_{nm}$ (we will say which we consider in a moment). Then we will show

$$C_{n}^{(m)} = \sum_{X \in \mathcal{X}_{nm}} N_{m}(X),$$

where $N_{m}(X)$ is the number of admissible $m$-labelings of $X$. As mentioned before, an object $X \in \mathcal{X}_{nm}$ only has the chance of a nonzero contribution if its level structure satisfies a obvious congruence condition: the number of elements in a nonzero level must be divisible by $m$. This condition is not sufficient, however, as one can see in Figure 4. Both trees have positive levels of even order, but there is no way to label the right tree admissibly: the parents of the vertices labeled $e$ are forced to have different labels.

![Figure 3. Two rooted trees $X, Y$ with their levels.](image)

![Figure 4. 2-labelings of $X$ and $Y$. The labeling of $Y$ is not admissible.](image)

4.7. We are now ready to give our combinatorial interpretations. For each we explain their level and hierarchical structures. Examples of all the objects are shown in Figures 6–7.
(i) **Plane trees.** The set $X_{nm}$ is the set of plane trees on $nm + 1$ vertices. The levels are the distance to the root, and the vertices are parents/children of each other if they are in the usual sense.

(ii) **Dyck paths.** The set $X_{nm}$ is the set of Dyck paths from $(0, 0)$ to $(2mn, 0)$. The elements of a path $P$ are its slabs, defined as follows. Let $R_P$ be the interior of the region bounded by $P$ and the $x$-axis. Then a slab is a connected component of the intersection of $R_P$ with an open strip $Y_k = \{(x, y) \in \mathbb{R}^2 \mid k - 1 < y < k\}$, where $k \geq 1$ is an integer. A slab is in level $k$ if it lies in $Y_k$; the zeroth level is empty. A slab $S$ is a parent of a slab $S'$ if $S$ sits in a lower level and sits under $S'$.

(iii) **Provisional ballot sequences.** Let $q$ be an indeterminate. The set $X_{nm}$ consists of sequences $B = (a_1, \ldots, a_{2nm})$ where $a_i \in \{\pm q^k \mid k \in \mathbb{Z}_{\geq 1}\}$ such that (i) $\sum_i a_i = 0$, (ii) if $a_i = q^k$ then $a_{i+1}$ must be either $q^{k+1}$ or $-q^k$, (iii) if $a_i = -q^k$ then $a_{i+1}$ must be either $-q^{k-1}$ or $q^k$, and (iv) $a_1 = q$. The elements of $B$ are its pairs: if $a_i = q^k$ then there is a smallest $j > i$ such that $a_j = -q^k$, and then $\{i, j\}$ form a pair. A pair is on level $k \geq 1$ if it equals $\{q^k, -q^k\}$; the zeroth level is empty. A pair $\{i, j\}$ is greater than a pair $\{l, m\}$ if $i < l < m < j$. Note that the level of a child is necessarily higher than that of its parent.

(iv) **Binary plane trees.** The set $X_{nm}$ consists of all binary plane trees with $nm$ vertices. The first level consists of the root and all the vertices that can be reached by moving right from the root. The $i$th level consists of all vertices that can be reached by moving left once from a vertex on level $i$, and then right an arbitrary number of times. Note that all levels are positive.

(v) **Triangulations.** Let $\Pi = \Pi_{nm+2}$ be a regular polygon with $nm + 2$ sides. Then $X_{nm}$ consists of triangulations $\Delta$ of $\Pi$ that do not have new vertices. The elements of $\Delta$ are its triangles, and the levels of $\Delta$ are given by the sets of left-turning triangles, which means the following. Fix once and for all an edge $e$ of $\Pi$ and let $T$ be the triangle of $\Delta$ meeting $e$. As one enters $T$ across $e$ there is a unique exiting edge $e_R$ to the right and one $e_L$ to the left. We say the triangle $T_R$ meeting $T$ at $e_R$, if there is one, is obtained by turning right, and the triangle $T_L$ across $e_L$, if there is one, is obtained by turning left. Then the first level of $\Delta$ consists of $T$ and all the triangles that can be reached from $T$ by turning left. The second level consists of the triangles that can be reached by turning right once from a triangle on the first level, and then turning left an arbitrary number of times. Continuing this process, each triangle in $\Delta$ gets placed into a unique positive level. The zeroth level is empty.

(vi) **Polygon gluings.** Let $\Pi = \Pi_{2mn}$ be a regular polygon with $2mn$ sides with a distinguished vertex. An oriented gluing of $\Pi$ is a partition of its sides into $n$ subsets of order $2m$, with the sides oriented such that as one moves clockwise around $\Pi$, the orientations in a given subset alternate (cf. the right column of Figure 7). We
Figure 5. The left and right triangles determined by an edge $e$, and a triangulation of a polygon with its level structure. The light grey triangles, which are obtained by turning left after entering across the edge $e$, are on level 1. The white triangles are on level 2, and the dark grey triangles are on level 3.

assume the first edges in each set are oriented such that clockwise is positive. The labeling of the edges induces an equivalence relation on the vertices of II, after one performs the identifications. Then $X_{nm}$ is the set of such gluings with the number of equivalence classes of vertices maximal. A vertex $b$ is a child of $a$ if the edge joining them is positive from $a$ to $b$. The distinguished vertex is at level 0, and the levels of the others are determined by requiring that passing from parent to child increases the level.

4.8. Figures 6–7 illustrate the combinatorial interpretations used in the computation of $C^{(2)}_2 = 6$. For this number each object has at most two levels; we show the first level using light grey and the second level using white. Labelings are indicated by the letters $a, b$. Note that only one object has more than one 2-labeling, namely the one appearing in the first three lines of each figure.

4.9. Theorem. The interpretations (i)–(vi) given in §4.7 count the numbers $C^{(m)}_n$, i.e. (6) holds.

Proof. First we claim that all the interpretations (i)–(vi) give the same counts. To see this, note that if $X_{nm}$ is any of the above sets (with no labelings), then $|X_{nm}| = C_{nm}$, and there are known bijections between the different objects (cf. [15, Theorem 1.5.1]). For the convenience of the reader, we recall these bijections. To simplify notation we set $m = 1$. 
Plane trees and Dyck paths. Let $T$ be a plane tree with $n + 1$ vertices. Then $T$ has $n$ edges, and one can build a length $2n$ Dyck path $\pi(T)$ as follows. One starts at the root of the tree and performs the preorder tree traversal: this is the traversal that visits a node, then recursively visits the left subtree in preorder traversal, then recursively visits the right subtree in preorder traversal. Along the way one crosses each edge of $T$ exactly twice, once going down, and once going up. For every move down one appends the step $(1, 1)$ to $\pi(T)$, and for every move up one appends $(1, -1)$. It is clear that the result is Dyck path; it has $2n$ steps, never goes below the $x$-axis, and ends at $(2n, 0)$. One can also easily reverse this process: given a Dyck path $\pi$, one builds a plane tree $T(\pi)$ whose preorder traversal is encoded by $\pi$.

Dyck paths and ballot sequences. Let $\pi$ be a Dyck path of length $2n$. We create a sequence $b(\pi) = (a_1, \ldots, a_{2n})$ with $a_i \in \{\pm 1\}$ by projecting onto the second coordinate: $(1, 1) \mapsto 1$ and $(1, -1) \mapsto -1$. Thus $b(\pi)$ records the change in height above the $x$-axis as one moves along $\pi$. Then $b(\pi)$ is clearly a ballot sequence: the condition that $\pi$ never goes below the $x$-axis ensures that the partial sums are nonnegative. It is also clear that this process can be reversed to give a map from ballot sequences to Dyck paths.
Plane trees and polygon gluings. Let $T$ be a plane tree. We can regard $T$ as embedded in the sphere $S^2$, for example by applying stereographic projection to the plane. If one cuts the sphere open along the edges of $T$, one obtains a polygon $\Pi_{2n}$ with $2n$ sides together with data of an oriented gluing: the edges come naturally in pairs, and the root becomes the distinguished vertex of $\Pi_{2n}$. To go backwards, if one starts with a polygon $\Pi_{2n}$ and identifies the edges in pairs, one obtains a topological surface $S$ together with an embedded connected graph $G$ ($G$ may have loops or multiple edges). The surface $S$ is orientable if and only if the gluing data
is oriented (as one goes around the boundary of the polygon, every pair of edges to be glued must appear with opposite orientations). If $S$ is orientable, then the graph $G$ is a tree if and only if $S$ is a sphere, which happens if and only if the number of equivalence classes of vertices is maximal after gluing; this follows from considering the Euler characteristic of $S$, computed as $1 - |\text{edges}| + |\text{vertices}|^2$.

**Triangulations and binary plane trees.** Let $\Delta$ be a triangulation of the polygon $\Pi_{n+2}$ with a distinguished edge $e$. One can make a binary plane tree $B(\Delta)$ by taking the dual $\Delta$ as follows (cf. Figure 8). The vertices of $B(\Delta)$ are the triangles of $\Delta$. Two vertices are joined by edges if and only if they correspond to adjacent triangles in $\Delta$. The distinguished edge $e$ sits in the boundary of one triangle, which determines the root of $B(\Delta)$. The resulting tree is embedded in the plane and is a binary tree exactly because each triangle has three sides. It is also easy to see that this construction is reversible.

**Plane trees and binary plane trees.** Finally we come to this bijection, due to de Bruijn–Morselt [4], which is the most interesting of all. We follow the presentation in [15, Theorem 1.5.1] closely.

Let $T$ be a plane tree on $n + 1$ vertices. We delete the root from $T$ and all edges incident to it. Then we remove all edges that are not the leftmost edge from any vertex. The remaining edges become the left edges in the binary tree $B(T)$. To find the right edges, we create edges horizontally across the vertices of $T$. In particular, given a vertex $v$ in $T$, we draw edges from each child $w$ of $v$ to the child of $v$ immediately to the right of $w$, if this child exists. Finally the root of $B(T)$ is the leftmost child of the root of $T$. (See Figure 9 for an example.) This process is easily seen to be reversible.

We now return to the original discussion. We claim that for $m > 1$, the same bijections prove that all interpretations give the same counts. One needs only to check that the definitions of admissible labelings are compatible. For instance, to pass from a binary tree $T$ to a Dyck path $\pi$, one starts at the root of the plane tree and traverses it in preorder (process the node, traverse the left subtree, then traverse the right subtree). As one descends $T$ along an edge, one steps up in $\pi$ by $(1, 1)$, and as one ascends along an edge, one steps down by $(1, -1)$. With this correspondence, it is evident that the levels of $T$ coincide with the slabs of $\pi$. Indeed, the slabs of $\pi$ are constructed so that the corresponding vertices of $T$ are at the same distance from the root, and the poset structure in the slabs mirrors that of the vertices of $[4, \text{§1.1}]$. There is a connection between polygon gluings and interpretation (59) in [15, Chapter 2]. This interpretation is that $C_n$ is the number of ways to draw $n$ nonintersecting chords joining $2n$ points on the circumference of a circle. If one starts with a connected oriented polygon gluing, and draws chords between the centers of the edge pairs, one obtains $n$ chords as in (59). The condition that these chords do not intersect is exactly equivalent to the resulting surface being a sphere.
Figure 8. Making a binary tree $B(\Delta)$ from a triangulation $\Delta$. The vertices of $B(\Delta)$ are white and its edges are dashed. The root vertex is circled.

Figure 9. Making a binary tree $B(T)$ from a plane tree $T$. We first delete the root of $T$ and all the edges that are not on the left. We create horizontal edges (dashed). The original (respectively, horizontal) edges become the left (resp., right) edges of $B(T)$.

It is also easy to see that the different notions of admissible labelings agree. The verifications for the other interpretations are similar.

To complete the proof, we must check that any one of them actually computes $C_n^{(m)}$. We will use admissibly labeled plane trees (i). First consider the case $m = 1$. Let $T$ be a (topological, i.e. non-embedded) tree on $n + 1$ vertices and choose a vertex $v$. A walk on $T$ contributing in (1) to $a_T(v)$ determines an embedding $i: T \to \mathbb{R}^2$ with $v$ as a root: one embeds $T$ with $v$ as root such that the preorder traversal follows the same steps as the walk. Conversely, it is clear that preorder traversal of a plane
tree with root \(v\) determines a walk on the underlying topological tree \(T\) contributing to \(a_T(v)\).

This correspondence between walks and embeddings does not induce a bijection, however. We claim two walks determine the same plane tree if and only if they are equivalent under an automorphism of \(T\). Indeed, let \(\pi, \pi'\) be two walks on \(T\) contributing to the computation of \(C_n\). Assume \(\pi\) begins at \(v\) and \(\pi'\) begins at \(v'\). If there is an automorphism of \(T\) carrying \(\pi\) to \(\pi'\), then it is clear that \(\pi, \pi'\) determine the same embedding \(i(T)\) of \(T\). On the other hand, suppose that \(\pi\) and \(\pi'\) determine the same plane tree \(i(T)\). Number the vertices of \(T\) arbitrarily with 1, 2, \ldots, \(n+1\). Use \(\pi\) and \(\pi'\) to construct two numberings \(\mu, \mu'\) of \(i(T)\). Then \(\mu, \mu'\) determine a permutation \(\sigma \in \text{Sym}(n+1)\) by comparison on \(i(T)\): if \(x \in i(T)\) receives \(j\) from \(\mu\) and \(j'\) from \(\mu'\), then \(\sigma(j) = j'\). This permutation \(\sigma\) determines a self-map of \(T\) that is evidently an automorphism. This proves

\[
\sum_{v \in T} \frac{a_T(v)}{|\Gamma(T)|}
\]

counts the number of rooted plane trees with underlying topological tree \(T\), which shows that (1) agrees with \(C_n\). This verifies Proposition 2.3, and proves our claim for \(m = 1\).

Next consider \(m > 1\). Let \(T \in \mathcal{T}_{n+1}\), choose a vertex \(v\), and label all other vertices using a set \(S\) of order \(n\). Then a walk on \(T\) beginning at \(v\) and contributing to \(a_T^{(m)}(v)\) determines a plane tree \(T'\) on \(nm+1\) vertices, again by requiring the preorder traversal of \(T'\) to match the walk. The tree \(T'\) has one vertex, the root, corresponding to \(v\), and has \(m\) vertices for each vertex in \(T\) other than \(v\). We can label the non-root vertices of \(T\) using the elements of \(S\), and it is clear that doing so produces an admissible \(m\)-labeling on \(T'\) with the same set \(S\) (Figure 10). To finish the proof, we note that a generalization of the argument above shows that two walks on \(T\) determine the same \(m\)-admissibly labeled plane tree (possibly with a permutation of the labels) if and only if they are equivalent under an automorphism of \(T\). This completes the proof.

\[\square\]

4.10. Remark. The interpretations in Theorem 4.9 make it possible to define various higher analogues of other standard numbers, such as Narayana numbers, and higher \(q\)-analogues. We have not pursued these definitions.

5. Generating functions and asymptotics

5.1. Let

\[
F_m(x) = \sum_{n \geq 0} C_n^{(m)} x^n
\]
be the ordinary generating function of the $C^{(m)}_n$. In this section we explain how to use the combinatorial interpretations in §4 to compute (7) to arbitrary precision as a power series in $x$. Then we will give a conjecture of the asymptotic behavior of $C^{(m)}_n$ that generalizes the famous formula

$$C_n \sim \frac{4^n}{n^{3/2} \sqrt{\pi}}, \quad (n \to \infty).$$

**5.2.** For any pair $d, r$, let $\lambda(r, d)$ be the dimension of the space of degree $d$ homogeneous polynomials in $r$ variables. We have

$$\lambda(r, d) = \binom{r - 1 + d}{r - 1}.$$

For any $k \geq 0$, let $W_m(k)$ be the number of partitions of a set of order $k$ into subsets of order $m$. Then $W_m(k) = 0$ unless $m \mid k$, and in this case we have

$$W_m(k) = \frac{k!}{(m!)^{k/m}(k/m)!}. $$

Put

$$\ell_m(x) = \sum_{r \geq 0} W_m(rm)\lambda(m, rm)x^r$$

and

$$h_m(x) = \sum_{r \geq 0} W_m(rm)x^r.$$

**5.3. Theorem.** Let $f_m(x)$ satisfy the functional equation

(8) \[ f_m(x) = x \ell_m(f_m(x)). \]
Then
\begin{equation}
F_m(x) = h_m(f_m(x)).
\end{equation}

Proof. We will prove that the coefficient of $x^n$ in $F_m(x)$ is $C_n^{(m)}$ by proving that it counts admissibly $m$-labeled plane trees with $nm + 1$ vertices. To do this, we need some standard results about constructing generating functions that enumerate trees satisfying various conditions. Our reference is [7], especially sections I.5, III.3, and III.5.

Let $\varphi(x) = \sum a_r x^r \in \mathbb{Z}_{\geq 0}[x]$, and suppose an ordinary generating function $f(x)$ satisfies the compositional relation
\begin{equation}
f(x) = x\varphi(f(x)).
\end{equation}

Then the coefficient $[x^n]f(x)$ counts plane trees on $n$ vertices such that (i) a vertex can have $r$ children if and only if $a_r \neq 0$, and (ii) each vertex with $r$ children can be assigned any one of $a_r$ colors [7, Proposition I.5]. For example, suppose $\varphi(x) = 1 + 3x^2$. Then $f(x) = x + 3x^3 + 18x^5 + 135x^7 + \cdots$ enumerates plane trees where each vertex is either a leaf or has 2 children, and each non-leaf can be painted three different colors [7].

Now consider $f_m(x)$. According to the above the functional equation (8) implies $f_m$ enumerates colored plane trees with colors determined by $\ell_m$. These “colors” can be interpreted as follows. Let $v$ be a vertex and assume $v$ has $r$ children. We break $v$ and each of its descendants into $m$ copies ordered by the planar embedding. We then (i) choose a partition of the $rm$ descendants into sets of order $m$ and (ii) choose edges from the $m$ copies of $v$ to the $rm$ descendants, so that the latter become children of the former (Figure 11). There are $W_m(rm)$ ways to accomplish (i) and $\lambda(m, rm)$ ways to accomplish (ii) (for (ii), we can think of each copy of $v$ as being indexed by a variable $x_i$, and then the configuration of edges determines a monomial of total degree $rm$).

The structures (i),(ii) are exactly the recursive structure on a plane tree with $nm + 1$ vertices determined by an admissible $m$-labeling, and in fact the coefficient of $x^n$ in $f_m$ almost equals $C_n^{(m)}$. The only problem is that the contribution of the root is wrong. According to the coefficients of $\ell_m$, in the computation of $f_m(x)$ we split the root, along with all the other vertices, into $m$ copies. This is not what we want to do. Indeed, the root should be treated differently: it should not be split into $m$ copies, although its $r$ descendants should, and the latter $rm$ points should be assigned a partition into subsets of order $m$. Thus the root should be receive $W_m(rm)$ colors. According to [7, III.5], in particular the discussion in Example III.8, to accomplish this we should consider
\begin{equation}
1 + W_m(m)f_m(x) + W_m(2m)f_m(x)^2 + \cdots.
\end{equation}

\[3\text{In fact the reader can check that }[x^{2n+1}]f(x) = 3^n \cdot C_n, \text{ although this is not needed in the sequel.}\]
This is exactly the definition of $F_m$ in (9), which completes the proof.

\[ \square \]

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{tree.png}
\caption{Building the recursive structure in an admissibly $m$-labeled plane tree. Let $m = 2$ and consider the coefficient $a_3 = W_2(6) \cdot \lambda(2, 6) = 15 \cdot 21$. Suppose a nonroot vertex $v$ has $r = 3$ children. We first split $v$ and its children into 2 copies. We then choose one of the 15 pairings of the 6 children and one of the 21 ways to attach the children back up to the parents (the indicated attachment corresponds to the monomial $x_1^2 x_2^3$). If $v$ is a root vertex with 3 children, then we don’t split $v$. We do split $v$’s children and pair them, but there is no choice in how the 6 children are attached back to $v$.}
\end{figure}

5.4. Remark. At this point the reader may be nostalgic for the recurrence relation satisfied by the classical Catalan numbers, which corresponds to the equation

\begin{equation}
xF_1^2 - F_1 + 1 = 0.
\end{equation}

For $m > 1$, the power series $F_m(x)$ does not appear to be algebraic, so unfortunately one does not have such a simple recurrence relation on its coefficients. However, experimentally one finds that there is an algebraic relation satisfied by $f_m$ and $F_m$:

\begin{equation}
f_m^2 - xF_m + x = 0.
\end{equation}

Relation (11) is easily checked when $m = 1$. We have $f_1 = xF_1$, so (11) is really the same as (10). For $m > 1$ a proof of (11) was given by Mark Wilson [16]. One observes that $\ell_m$ and $h_m$ are related by

\[ x\ell_m(x) = h_m(x) - 1, \]

from which (11) easily follows.

The relation (11) gives a connection between pairs of objects computing hypergraph Catalan numbers in the spirit of that encoded by (10). Consider $m = 2$. We have

\[ f_2(x) = x + 3x^2 + 24x^3 + 267x^4 + \cdots, \]
and
\[ f_2(x)^2 = x^2 + 6x^3 + 57x^4 + 678x^5 + \cdots = x(F_2(x) - 1)). \]

We see \( C^{(2)}_3 = 57 \) as the coefficient of \( x^4 \) in \( f_2(x)^2 \), which comes from the coefficients of \( x, x^2, x^3 \) in \( f_2(x) \) via

\[ 57 = 24 \cdot 1 + 3 \cdot 3 + 1 \cdot 24. \]

This is certainly reminiscent of the classical Catalan relation, although there is an important difference. The numbers involved on the right of (12) are connected with \( C^{(2)}_2, C^{(2)}_3, C^{(2)}_4 \), but they enumerate proper subsets of the associated objects, not the full sets. Indeed, this is obviously true, since both sides of (12) involve the same number \( C^{(2)}_3 \). We have not attempted to explore this connection further.

5.5. We conclude this section by discussing asymptotics for the \( C^{(m)}_n \). The results here are purely experimental; none have been proved, although based on our numerical experiments we are very confident in them. We learned our basic experimental technique from Don Zagier, who calls it multiplying by \( n^8 \); an excellent lecture by him at ICTP demonstrating the method can be found online [17].

Suppose one has a sequence \( a = \{a_n\}_{n \geq 0} \) that one believes satisfies an asymptotic of the form

\[ a_n \sim C_0 + C_1/n + C_2/n^2 + \cdots \quad (n \to \infty). \]

It may be difficult to extract \( C_0 \) even when \( n \) is large, since \( C_1/n \) might still be non-negligible. However it is possible to wash away the contributions of the nonconstant terms \( C_k/n^k, k \geq 1 \). One multiplies both sides of (13) by \( n^8 \) (or any other reasonable even power of \( n \)), and then applies the difference operator \((8!)^{-1} \Delta^8\) to both sides, where \( \Delta a \) is the sequence \( (\Delta a)_n := a_{n+1} - a_n \). The operator \( \Delta^k \) annihilates any polynomial \( p(n) \) of degree \( < k \), takes \( n^k \) to \( k! \), and takes \( n^{-l} \) to a rational function in \( n \) of homogeneous degree \( -l - k \). Thus applying \((8!)^{-1} \Delta^8\) to both sides of (13) yields

\[ \frac{1}{8!}(\Delta^8 n^8 a)_n \sim C_0 + C_9 p_{-9}(n) + C_{10} p_{-10}(n) + \cdots, \]

where \( p_{-k}(n) \) denotes a rational function in \( n \) of homogeneous degree \( -k \). If one then evaluates the left of (14) at a large value of \( n \), the effects of \( C_k, k \geq 9 \) are negligible on the right and one clearly sees \( C_0 \). One can then repeat the process with the sequence \( n(a_n - C_0) \) to find \( C_1 \), and so on.

5.6. We illustrate with the series \( F_2(x) = 1 + x + 6x^2 + 57x^3 + 678x^4 + \cdots \). Playing with the data one makes the ansatz

\[ C^{(2)}_n \sim KA^n n! n^\rho \]

for some constants \( K, A, \rho \). Indeed, apart from the \( n! \), the right of (15) is typical for these kinds of problems, and was our initial guess; it quickly became evident that \( n! \)
needed to be included. The series

\[ a_n := C_n^{(2)}/n! \]

should then grow exponentially, and the sequence of ratios \( b_n = a_n/a_{n-1} \) should satisfy (13) with \( C_0 = A \). Indeed, using 100 terms of \( F_2 \) and \( \Delta^{16} \) we get

\[ A \approx 2.000000000068961809 \ldots . \]

Next we consider the sequence

\[ a_n := C_n^{(2)}/(2^n n!), \]

which we expect to be asymptotic to \( Kn^\rho \). We can detect \( \rho \) using the sequence

\[ b_n = \left( \Delta \log a_n \right) n/(\log(n+1) - \log n), \]

which satisfies (13) with \( C_0 = -\rho \). Again with 100 terms and \( \Delta^{16} \) we get

\[ -\rho \approx 0.4999999726715 \ldots . \]

Hence we have

\[ C_n^{(2)} \sim K \cdot \frac{2^n n!}{\sqrt{n}} \]

and we must find the constant \( K \). We consider \( b_n = C_n^{(2)} \sqrt{n}/2^n n! \) and look for \( C_0 \). This time \( K \) is more difficult to find. Taking 200 terms and applying \( \Delta^{16} \), this number appears to be 5.05704458036912766 \ldots . We use the Inverse Symbol Calculator [2], which attempts to symbolically reconstruct a given real number using various techniques, and find

\[ K \approx 2e^{3/2}/\sqrt{\pi}. \]

(The \( e^{3/2} \) is surprising, but is apparently correct. Using 600 terms of the sequence, we can verify that \( K\sqrt{\pi}/2 \) agrees with \( e^{3/2} \) with relative error < \( 10^{-76} \).) The conclusion is

\[ C_n^{(2)} \sim e^{3/2} \cdot \frac{2^{n+1} n!}{\sqrt{\pi n}}, \quad (n \to \infty). \]

We present asymptotics for the \( C_n^{(m)} \) as a conjecture:

**5.7. Conjecture.** Let \( m > 1 \). Then as \( n \to \infty \), we have

\[ C_n^{(m)} \sim K_m \cdot \frac{(m-1)!^{n+1} (n!)^{m-1}}{(\pi n)^{(m-1)/2}}, \]

where the constant \( K_m \) is defined by

\[
K_m = \begin{cases} 
  e^{3/2} & \text{if } m = 2, \\
  2^{(m-1)/2} \frac{3}{2} \frac{5}{2} \cdots \frac{m-1}{2} / m^{(2m-3)/2} & \text{if } m \geq 3 \text{ and is odd, and} \\
  \sqrt{2}^{(m-1)/2} \frac{3}{2} \frac{5}{2} \cdots \frac{m-1}{2} / m^{(2m-3)/2} & \text{if } m \geq 4 \text{ and is even.}
\end{cases}
\]
5.8. Remark. We have tested Conjecture 5.7 numerically using 100 terms of \( F_m(x) \) for all \( m \leq 30 \). We have not systematically tried to find higher terms in the asymptotic expansion of \( C_n^{(m)} \), as in (13).

6. Connection with matrix models

6.1. We finish by explaining how the numbers \( C_n^{(m)} \) are related to hypergraphs and matrix models. We first explain the connection between graphs, matrix models, and the usual Catalan numbers. For more information, we refer to Harer–Zagier [8], Etingof [5, §4], Lando–Zvonkin [9], and Eynard [6].

6.2. Let \( d\mu_2(x) \) be the measure on polynomial functions on \( \mathbb{R} \) with moments

\[
\langle x^r \rangle_2 := \int_{\mathbb{R}} x^r d\mu_2(x) = W_2(r),
\]

where \( W_2(r) \) is the number of pairings on a set of order \( r \). It is well known that \( d\mu_2(x) \) is essentially the Gaussian measure, up to normalization: we have

\[
\langle x^r \rangle_2 = (2\pi)^{-1/2} \int_{\mathbb{R}} x^r e^{-x^2/2} dx,
\]

where \( dx \) is the usual Lebesgue measure on \( \mathbb{R} \).

Let \( g_1, g_2, \ldots \) be a family of indeterminates, and let \( S(x) \) be the formal power series \( \sum_{r \geq 1} g_r x^r / r! \). We can compute the expectation \( \langle \exp(S(tx)) \rangle_2 \) as a formal power series in \( t \) with coefficients in the polynomial ring \( \mathbb{Q}[g_1, g_2, \ldots] \). We have

\[
\langle \exp(S(tx)) \rangle_2 = 1 + A_2 t^2 / 2 + A_4 t^4 / 8 + A_6 t^6 / 48 + \cdots
\]

where

\[
A_2 = g_1^2 + g_2, \quad A_4 = g_1^4 + 6g_1^2 g_2 + 4g_1 g_3 + 3g_2^2 + g_4,
\]

\[
A_6 = g_1^6 + 15g_1^4 g_2 + 20g_1^3 g_3 + 45g_1^2 g_2^2 + 15g_2^3 g_4 + 60g_1 g_2 g_3 + 6g_1 g_5 + 15g_3^2 + 15g_2 g_4 + 10g_3^2 + g_6.
\]

The series (16) can be interpreted as a generating function for graphs weighted by the inverse of the orders of their automorphism groups. Let \( \mathbf{n} = (n_1, n_2, \ldots) \) be a vector of nonnegative integers, with \( n_i \) nonzero only for finitely many \( i \). Let \( |\mathbf{n}| = \sum n_i \). We say a graph \( \gamma \) has profile \( \mathbf{n} \) if it has \( n_i \) vertices of degree \( i \). Let \( G(\mathbf{n}) \) be the set of all graphs of profile \( \mathbf{n} \), up to isomorphism (we allow loops and multiple edges). By an automorphism of a graph, we mean a self-map that permutes edges and vertices. In particular, automorphisms include flipping loops and permuting multiedges. For any \( \gamma \in G(\mathbf{n}) \), let \( \Gamma(\gamma) \) be its automorphism group. Then we have

\[
\langle \exp S(tx) \rangle_2 = \sum_{\mathbf{n}} t^{|\mathbf{n}|} \sum_{\gamma \in G(\mathbf{n})} \prod_{i} g_i^{n_i} |\Gamma(\gamma)|.
\]
For an example, consider the term $5g_3^2/24$. There are two graphs with this profile, shown in Figure 12. The left has $2 \cdot 2 \cdot 2$ automorphisms, and the right has $2 \cdot 3!$, which gives $1/8 + 1/12 = 5/24$.

**Figure 12.** The two graphs with profile $g_3^2$.

6.3. Now we want to replace the Gaussian measure, which is connected to counting pairings of a set, with something that is connected to the numbers $W_{2m}(r)$. Let $d\mu_{2m}(x)$ be the “measure” on polynomial functions on $\mathbb{R}$ that gives the monomial $x^r$ the expectation $W_{2m}(r)$. More precisely, we consider the function taking $x^r$ to $W_{2m}(r)$ and extend linearly to polynomials. This is not a measure in the usual sense, although formally we can regard it as such. The “expectation” $\langle \exp S(tx) \rangle_{2m}$ is then a well-defined power series in $t$, and has a combinatorial interpretation via hyperbaggraphs.

Recall that a hypergraph, a notion due to Berge [1], on a vertex set $V$ is a collection of subsets of $V$, called the hyperedges. The degree of a vertex is the number of hyperedges it belongs to, and a hypergraph is regular if these numbers are the same for all vertices. The order of a hyperedge is its number of vertices. If all hyperedges have the same order, we say that the hypergraph is uniform.

Now suppose we allow $V$ to be a multiset, in other words a set with a multiplicity map $V \rightarrow \mathbb{Z}_{\geq 1}$. Then these constructions lead to hyperbaggraphs, due to Ouvrard–Le Goff–Marchand-Maillet [11]. We extend the notions of regularity and uniformity above by incorporating the multiplicity in an obvious way (the order of a subset of a multiset is sum of the multiplicities of its elements).

With these definitions, the expectation $\langle \exp S(tx) \rangle_{2m}$ now enumerates uniform hyperbaggraphs of all profiles weighted by the inverses of their automorphism groups, where each hyperedge has $2m$ elements. For example,

$$\langle \exp S(tx) \rangle_4 = 1 + B_4 t^4/24 + B_8 t^8/1152 + \cdots$$

where

$$B_4 = g_4^1 + 6g_2^1g_2 + 4g_1g_3 + 3g_2^2 + g_4, \quad B_8 = g_8^1 + 28g_4^1g_2 + 56g_5g_3 + \cdots + 35g_4^2 + g_8.$$  

The computation of the contribution $35g_4^2/1152$ is as follows. There are three hyperbaggraphs of this profile, each with two hyperedges. The underlying set of vertices has 2 elements $a, b$, and we represent a hyperedge by a monomial in these variables.

---

4In the CS literature, multisets are sometimes called bags.
The profile $g_4^2$ means that each vertex has degree 4, and since $2m = 4$ we must have uniformity 4. Thus we want pairs of monomials in $a, b$ of total degree 4. This gives

\[(18) \{a^4, b^4\}, \{a^3b, ab^3\}, \{a^2b^2, a^2b^2\}.\]

The orders of the automorphism groups are

\[(19) 2 \cdot (4!)^2, 2 \cdot (3!)^2, 2 \cdot 2 \cdot (2!)^2 (2!)^2.\]

For example, the automorphisms of the last hyperbagraph come to interchanging the vertices, interchanging the two hyperedges, and the internal flips within the hyperedges; the last type of automorphism cannot occur for graphs. Adding the inverses of these orders, one finds $1/1152 + 1/72 + 1/64 = 35/1152$, which agrees with $B_8$ above.

As a final remark, we note that $A_4 = B_4$. This is a general phenomenon: one can show that the coefficient of $t^n$ in $\langle \exp S(tx) \rangle_{2m}$ is the complete exponential Bell polynomial $Y_n(g_1, \ldots, g_n)$, divided by $(2m)!^d$, where $d = n/2m$. We refer to [3, p. 134, eqn. 3b] for the definition of these; the coefficients of the $Y_n$ can be found on OEIS as A178867.

6.4. Now we pass to matrix models. Let $V = V_N$ be the real vector space of $N \times N$ complex Hermitian matrices. The space $V$ has real dimension $N^2$. For any polynomial function $f : V \to \mathbb{R}$, define

\[(20) \langle f \rangle = C^{-1} \int_V f(X) \exp(-\text{Tr}X^2/2) dX,\]

where $\text{Tr}(X) = \sum_i X_{ii}$ is the sum of diagonal entries and the constant $C$ is determined by the normalization $\langle 1 \rangle = 1$. The measure $\exp(-\text{Tr}X^2/2) dX$ is essentially the product of the measures $d\mu_2(x)$ from §6.2 taken over the real coordinates of $V$. The only difference is that for any off-diagonal entry $Z_{ij} = X_{ij} + \sqrt{-1}Y_{ij}$, we have rescaled the measure so that for even $r$ we have $\langle X_{ij}^r \rangle_2 = \langle Y_{ij}^r \rangle_2 = W_2(r)/2^{r/2}$.

Now consider (20) evaluated on the polynomial given by taking the trace of the $r$th power:

\[(21) P(N, r) = \langle \text{Tr}X^r \rangle .\]

For $r$ odd (21) vanishes for all $N$. On the other hand, for $r$ even and $N$ fixed, it turns out that $P(N, r)$ is an integer, and as a function of $N$ is a polynomial of degree $r/2 + 1$ with integral coefficients.

Furthermore, the number $P(N, r)$ has the following remarkable combinatorial interpretation. Let $\Pi_r$ be a polygon with $r$ sides. Any pairing $\pi$ of the sides of $\Pi_r$ determines a topological surface $\Sigma(\pi)$ endowed with an embedded graph (the images of the edges and vertices of $\Pi_r$). Let $v(\pi)$ be the number of vertices in this embedded
graph. Then we have

\[ P(N, r) = \sum_{\pi} N^{v(\pi)}, \]

where the sum is taken over all oriented pairings of the edges of \( \Pi_r \) such that the resulting topological surface \( \Sigma_{\pi} \) is orientable. For example, we have

\[ P(N, 0) = N, P(N, 2) = N^2, P(N, 4) = 2N^3 + N, \]
\[ P(N, 6) = 5N^4 + 10N^2, \quad P(N, 8) = 14N^5 + 70N^3 + 21N. \]

The pairings yielding \( P(N, 4) \) are shown in Figure 13.

![Figure 13. Computing \( P(N, 4) = 2N^3 + N \).](image)

6.5. One can see from (23) that the leading coefficient of \( P(N, r) \) is none other than the Catalan number \( C_{r/2} \). This follows easily from the interpretation of the Catalan numbers in terms of polygon gluings ((vi) from §4.7). Indeed, the leading coefficient counts the number of oriented pairings of \( \Pi_r \) such that the number of vertices \( v(\pi) \) in the orientable surface \( \Sigma_r \) is maximal; this is exactly the interpretation above.

6.6. Now we modify the matrix model. We replace the Gaussian measure

\[ \exp(-\text{Tr} X^2/2) dX \]

by the product of the formal measures \( d\mu_{2m}(x) \) taken over the real coordinates; again we rescale on the off-diagonal coordinates so that for \( r \equiv 0 \mod 2m \) we have

\[ \langle X_{ij}^r \rangle_{2m} = \langle Y_{ij}^r \rangle_{2m} = W_{2m}(r)/2^r/2m. \]

We write the corresponding formal measure by \( d\mu_{2m}(X) \). Then we obtain a new matrix model where polygons are glued by grouping their edges into subsets of order \( 2m \) instead of pairs. As above one can see that for \( r \equiv 0 \mod 2m \), the integrals

\[ \int_V \text{Tr} X^r d\mu_{2m}(X) \]

are polynomials \( P_{2m}(N, r) \) in the dimension \( N \). For example, when \( 2m = 4 \) we have

\[ P_4(N, 4) = N^2, \quad P_4(N, 8) = 6N^3 + 21N^2 + 8N, \]
\[ P_4(N, 12) = 57N^4 + 715N^3 + 2991N^2 + 2012N. \]
The hypergraph Catalan numbers $C^{(m)}_r$ are the leading coefficients of these polynomials. More precisely, we have

$$P_{2m}(N,2mr) = C^{(m)}_r N^{r+1} + \cdots.$$  

A direct computation with the definition \cite{24} shows that these numbers are computed via Definition \cite{25}. More details about these matrix models will appear in a future publication.

**References**


**Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003–9305**

*E-mail address: gunnells@math.umass.edu*