1 Species of structures

A combinatorial species assigns to each finite set $U$ a corresponding finite set of combinatorial objects built from $U$. The set $U$ is called the label set. In statistical mechanics it is a set of particles.

Example: The power set species $P$ assigns to $U$ the set $P[U]$ of all subsets of $U$.

Example: The partition species $\text{Part}$ assigns to $U$ the set $\text{Part}[U]$ of partitions of $U$ (into non-empty subsets).

Example: The set indicator species $E$ assigns to $U$ a set with a single point (for instance, it can be taken to be $\{U\}$).

Example: The one-point set indicator species $E_1$ assigns to a one-point set $U = \{j\}$ a set with a single point; it assigns to every other set the empty set.

In general, if $\#U = n$, then we write $f_n = #F[U]$. This is the number of combinatorial objects. There is an associated exponential generating function

$$F(z) = \sum_{n=0}^{\infty} \frac{1}{n!} f_n z^n.$$  (1)

These functions are combinatorial analogs of functions of the activity parameter $z$ for a particle system in statistical mechanics. (The activity tunes the density, much as the temperature tunes the energy.)

Example: For the power set species $p_n = 2^n$, so $P(z) = e^{2z}$.

Example: For the partition species $\text{part}(z) = B_n$, the $n$th Bell number.

These numbers are perhaps not so familiar, but the exponential generating function works out to be $\text{Part}(z) = e^{e^z-1}$. See below for a derivation.

Example: For the set indicator species $E(z) = e^z$.

Example: For the one-point set indicator species $E(z) = z$.

2 Species operations

There are operations on species. Here are four.

$\bullet$ $(F + G)[U] = F[U] + G[U]$
\( (F \ast G)[U] = \sum_{V \subseteq U} F[U] \times G[U \setminus V] \)

\( (F \circ G) = \sum_{\Gamma \in \text{Part}[U]} F(\Gamma) \times \prod_{V \in \Gamma} G[V] \)

\( F'[U] = F[U + \{\ast\}] \)

There are corresponding operators on the exponential generating functions.

\( (F + G)(z) = F(z) + G(z) \)

\( (F \ast G)(z) = F(z)G(z) \)

\( (F \circ G)(z) = F(G(z)) \)

\( F'(z) = \frac{d}{dz} F(z) \)

Example: It is easy to see that \( P = E \ast E \). This gives another explanation of the fact that \( P(z) = e^z e^z = e^{2z} \).

Example: It is easy to see that \( \text{Part} = E \circ E_+ \), where \( E_+[U] \) is a single point when \( U \) is non-empty, otherwise empty. Since \( E_+(z) = e^z - 1 \), it follows that \( \text{Part}(z) = e^{e^z-1} \).

If \( F \) is a species, then \( F^\bullet \) is the species of rooted structures. Thus \( F^\bullet[U] \) consists of all ordered pairs consisting of a structure in \( F[U] \) and a point in \( U \).

It is clear that \( f_n^\bullet = nf_n \). On the level of exponential generating functions we have

\[ F^\bullet(z) = z \frac{d}{dz} F(z). \quad (2) \]

### 3 Connected graphs

Let \( U \) be a finite set. Each element of \( U \) is a vertex. A graph consists of a set of two-element subsets of \( U \). Each such two-element subset is called an edge.

Then \( G[U] \) is the set of graphs with vertex set \( U \).

Suppose \( \#U = n \). Then

\[ g_n = \#G[U] = 2^\binom{n}{2}. \quad (3) \]

A partition of a set \( U \) is a collection \( \Gamma \) of disjoint non-empty subsets of \( U \) whose union is \( U \). Each set \( V \) in the partition \( \Gamma \) is called a part of the partition. Let \( \text{Part}[U] \) denote the set of all partitions of \( U \).

Let \( \Gamma \) be a partition of \( U \). If \( G \) is a graph on \( U \), then \( \Gamma \) partitions \( G \) if for every edge \( \ell \) in \( G \) there is a \( V \) in \( \Gamma \) such that \( \ell \) is a subset of \( V \). A graph on \( U \neq \emptyset \) is connected if it has no non-trivial partition. Let \( C[U] \) denote the set of connected graphs on \( U \). If \( U \) has one point, then the graph with no edges is connected. If \( U \) has two points, then there is just one connected graph. If \( U \) has three points, then there are four connected graphs.

If \( G \) is a graph on \( U \), then there exists a partition \( \Gamma \) of \( U \) such that on each part \( V \) the edges of \( G \) that are subsets of \( V \) form a connected graph. In other words,

\[ G = \bigsqcup_{V \in \Gamma} G_V, \quad (4) \]
where each $G_V$ is a connected graph on $V$. This equation says that the species of graphs is the combinatorial exponential of the species of connected graphs.

$$ G = E \circ C. \hfill (5) $$

In other words,

$$ G[U] = \sum_{r \in \text{Part}[U]} \prod_{V \in r} G[V]. \hfill (6) $$

This says that for every graph there is a partition of the vertices with connected graphs on each part. For the exponential generating functions it follows that

$$ G(z) = e^{C(z)}. \hfill (7) $$

This equation relates the rather mysterious connected graph species to the well-understood graph species. The analog in statistical mechanics is the equation that writes the grand partition function as the exponential of the pressure.

Let $C^*[U]$ be the set of all vertex-rooted connected graphs. The first Mayer equation is

$$ G^* = C^* \star G. \hfill (8) $$

This says that

$$ G^*[U] = \sum_{V \subseteq U} C^*[V] \times G[U \setminus V]. \hfill (9) $$

That is, a rooted graph on a vertex set $U$ defines a rooted connected graph on a subset $V$ (the subset with the root in it), together with a graph on the complement. There is a corresponding exponential generating function relation

$$ G^*(z) = C^*(z) G(z). \hfill (10) $$

Here is a physics dictionary. The combinatorial objects index the terms in the expansions of the functions in powers of the activity.

- Graphs — grand partition function — $G(z)$
- Connected graphs — pressure — $C(z)$
- Rooted connected graphs — density — $C^*(z)$

4 Trees

A tree is a minimal connected graph. Let $T[U]$ denote the set of trees on $U$. Cayley proved that $t_n = n^{n-2}$. If $U$ has one point, then the graph with no edges is a tree. If $U$ has two points, then there is just one tree. If $U$ has three points, then there are three trees.

A rooted tree (or vertex-rooted tree) is an ordered pair consisting of a tree on $U$ and a point $j$ in $U$. Let $T^*[U]$ denote the set of rooted trees on $U$. It is clear that $t_n^* = mn^{n-2} = n^{n-1}$.3
There is a fixed point equation for $T^\bullet$. This is

$$T^\bullet = E_1 \ast (E \circ T^\bullet).$$  \hspace{1cm} (11)

Here $(E_1 \ast F)[U] = \sum_{j \in U} F[U \setminus \{j\}]$. So this equation says that

$$T^\bullet[U] = \sum_{j \in U} \sum_{T \in \text{Part}[U \setminus \{j\}]} \prod_{V \in T} T^\bullet[V].$$  \hspace{1cm} (12)

A rooted tree consists of a root point together with a partition of the other vertices, on each part of which there is a rooted tree. The exponential generating function relation is

$$T^\bullet(z) = ze^{T^\bullet(z)}.$$  \hspace{1cm} (13)

This equation may be solved to derive Cayley’s theorem.

There is no obvious fixed point equation for $T$. However there is an identity relating $T$ with $T^\bullet$. This is the dissymmetry theorem.

$$T + T^\bullet + T^- = T^\bullet + T^-.$$  \hspace{1cm} (14)

Here $T^\bullet$ produces vertex-rooted trees, as before. $T^-$ produces edge-rooted trees. Finally, $T^{\bullet-}$ produces trees that are rooted at some adjacent vertex-edge pair.

Say that an “element” of a tree is either a vertex or an edge. Every tree has a central element.

The right hand side $T^\bullet + T^-$ of the equation is the set of trees that are rooted at some element. One can think of $T$ as the set of trees that are rooted at the central element. Thus it is sufficient to show that $T^{\bullet-}$ corresponds to non-central elements.

Given a non-central element, then there is an adjacent element pair consisting of the element together with the next element closest to the center. Going the other way, given an adjacent element pair, there is a non-central element consisting of the element farthest from the center. This completes the proof of the dissymmetry theorem.

Now we can look at the extra terms $T^-$ and $T^{\bullet-}$. Giving an element of $T^-[U]$ is giving a partition of $U$ into two subsets, together with a vertex-rooted tree on each part. (The distinguished edge connects the roots.) Let $E_2$ be the species that returns a single point for every 2-element set, otherwise the empty set. This is the two-point indicator species. Then

$$T^- = E_2 \circ T^\bullet.$$  \hspace{1cm} (15)

On the other hand, giving an element of $T^{\bullet-}[U]$ is giving a splitting (ordered partition) into two subsets, together with a vertex-rooted tree on each part. (The distinguished edge connects the roots.) So these constructions are also related to vertex-rooted trees. In fact,

$$T^{\bullet-} = T^\bullet \ast T^\bullet.$$  \hspace{1cm} (16)
Putting this all together, we get the combinatorial identity
\[ T + T^* \ast T^* = T^* + E_2 \circ T^*. \] (17)

On the level of exponential generating functions we obtain
\[ T(z) + T^*(z)T^*(z) = T^*(z) + \frac{1}{2}(T^*(z))^2. \] (18)

This remarkable identity relates trees to rooted trees. It can also be written as
\[ T(z) = T^*(z) - \frac{1}{2}(T^*(z))^2, \] (19)
but then the combinatorial origin is obscured.

5 2-vertex connected graphs

Consider a connected graph $G$ on $U$. Let $j$ be a vertex in $U$. Let $G_j$ be the graph on $U \setminus \{j\}$ obtained by removing the vertex $j$ and all edges from $j$. Then $G$ is 2-vertex connected if for each $j$ in $U$ the graph $G_j$ is connected. Let $\mathcal{B}[U]$ be the set of all 2-vertex connected graphs on $U$. If $U$ has one point, then there are no 2-vertex connected graphs. If $U$ has two points, then there is one 2-vertex connected graph. If $U$ has three points, then there is one 2-vertex connected graph.

If $G$ is a connected graph, then a vertex $j$ such that $G_j$ is not connected is called a cut-point. A maximal 2-vertex connected graph on a subset $V \subset U$ is called a block. Every connected graph $G$ on $U$ may be decomposed into blocks on subsets $V$ joined at cut-points. The vertices of the blocks may overlap at cut-points, but the blocks partition the edges.

As a first exercise, let $C^\circ$ be the species of block-rooted connected graphs. Then
\[ C^\circ = \mathcal{B} \circ C^*. \] (20)

The left hand side is an ordered pair consisting of a connected graph and a block. Each point of the block defines a rooted connected graph on a subset, corresponding to the vertices that are connected to to the block through this point. These subsets partition the vertex set of the original graph. The block defines a 2-vertex connected graph on the partition (that is, on the roots of the connected graphs).

The second Mayer equation is
\[ C^* = E_1 \ast (E \circ \mathcal{B}' \circ C^*). \] (21)

Here $\mathcal{B}'[U] = \mathcal{B}[U + \{\ast\}]$. The corresponding exponential generating function equation is
\[ C^*(z) = ze^{\mathcal{B}'(C^*(z))}. \] (22)

This marvelous fixed point equation would be even better if one had a good handle on $\mathcal{B}(z)$ and $\mathcal{B}'(z)$. 

5
To understand the combinatorial equation, start with a rooted connected graph $G$ on $U$ with root at $j$. Remove the root and the edges from the root, forming a graph $G_j$. This graph $G_j$ breaks up into connected graphs. This gives the first partitioning, that of $U \setminus \{j\}$.

Consider one of the connected graphs $H$ on a set of vertices $W$. There is a non-empty subset of vertices in $W$ that are directly connected to $j$. These vertices together with $j$ are 2-vertex connected, so they belong to a block on $W$ together with $j$. The vertices of this block in $W$ are each the root of a connected graph on some subset of $W$. This gives a second partitioning, of $W$ into parts with rooted connected graphs. So we have a $(B' \circ C^*)[W]$ structure. This consists of a partition of $W$, a connected graph on each set of the partition, and a $B'$ structure on the parts of the partition.

Now we turn to connected graphs. Given a connected graph, there is a tree with white and black vertices associated with it. The white vertices are the blocks of the graph, while the black vertices are the cut-points of the graph. The edges of the tree connect white vertices with black vertices, that is, they connect blocks with cut-points.

The end points of the tree are white vertices. The center of such a bi-colored tree is always a vertex. Thus for every connected graph there is a center block or cut-point.

The vertex dissymmetry theorem for connected graphs states that

$$C + C^{\bullet \circ} = C^\bullet + C^\circ. \quad (23)$$

Here $C^{\bullet \circ}$ gives connected graphs with designated block and designated vertex in the block.

The proof is almost as before. Define an element of a connected graph to be either a vertex or a block. The right hand side consists of the connected graphs with a designated element. The connected graphs on the left correspond to the situation where the designated element is the center (and is hence a cut-point or a block). So we need to consider the elements that are not the center.

For each such element there is an element of the other kind that is closer to the center. For a block there is a closer cut-point. For a vertex there is the block to which it belongs that is closest to the center.

Going backward, start with a pair of elements consisting of a block and a vertex in the block. If the vertex is a cut-point, then distinguish the vertex or the block farther from the center. If the vertex is not a cut-point, it is not the center, and one can distinguish the vertex.

Here is an expression for $C^{\bullet \circ}$. It is

$$C^{\bullet \circ} = C^\bullet * (B' \circ C^*). \quad (24)$$

This says that there is an ordered partition with a rooted connected graph on the first part and a $B' \circ C^*$ structure on the second part. This structure consists of a partition of the second part into rooted connected graphs, plus a 2-connected graph on these roots together with the root of the first connected graph.
The conclusion is that
\[ \mathcal{C} + \mathcal{C}^* \ast (B' \circ \mathcal{C}^*) = \mathcal{C}^* + B \circ \mathcal{C}^*. \] (25)

The corresponding exponential generating function identity is
\[ \mathcal{C}(z) + \mathcal{C}^*(z)B'(\mathcal{C}^*(z)) = \mathcal{C}^*(z) + \mathcal{B}(\mathcal{C}^*(z)). \] (26)

This expresses the pressure \( \mathcal{C}(z) \) in terms of the density \( \mathcal{C}^*(z) \). This is a form of the virial expansion.

6 2-edge connected graphs

There is a similar theory of 2-edge connected graphs. This theory seems not to be relevant to statistical mechanics, but something like it is important in the context of Feynman diagrams.

A graph is 2-edge connected if it cannot be disconnected by removing an edge. If \( U \) has one point, then it is regarded as being 2-edge connected. If \( U \) has two points, then there are no 2-edge connected graphs. If \( U \) has three points, then there is one 2-edge connected graph.

A maximal 2-edge connected subgraph is called a lump. A connected graph on \( U \) defines a partition of the vertex set in which there is a lump on each part. An edge that connects two lumps is called a bridge. The edge set is partitioned into the disjoint union over lumps and bridges of the edge sets associated with the individual lumps and individual bridges.

A connected graph defines a tree, in which the lumps are the vertices, and the edges are the bridges.

Let \( \mathcal{M} \) be the species of 2-edge connected graphs. Let \( \mathcal{C}^\leftrightarrow \) be the species of connected graphs with a designated bridge. Then
\[ \mathcal{C}^\leftrightarrow = E_2 \circ \mathcal{C}^*. \] (27)

Let \( \mathcal{C}^\circ \) be the species of connected graphs with a designated lump. Then
\[ \mathcal{C}^\circ = \mathcal{M} \circ (E_1 \ast (E \circ \mathcal{C}^*)). \] (28)

Let \( \mathcal{C}^{\circ \leftrightarrow} \) be the species of connected graphs with a designated lump-bridge pair. The edge dissymmetry theorem for connected graphs says that
\[ \mathcal{C} + \mathcal{C}^{\circ \leftrightarrow} = \mathcal{C}^\circ + \mathcal{C}^{\leftrightarrow}. \] (29)

Explicitly, this says that
\[ \mathcal{C} + \mathcal{C}^* \ast \mathcal{C}^* = \mathcal{M} \circ (E_1 \ast (E \circ \mathcal{C}^*)) + E_2 \circ \mathcal{C}^*. \] (30)

There is also a fixed point equation
\[ \mathcal{C}^* = \mathcal{M}^* \circ (E_1 \ast (E \circ \mathcal{C}^*)). \] (31)

The exponential generating function equations follow in a routine way.