Techniques for counting the structural isomers of alkanes and monosubstituted alkanes

Séamus Delaney  Mourtadha Badiane  Cian O’Brien
Michael Flattery  Brian Dineen

2nd Annual Stokes Modelling Workshop

June 18, 2015
The problem our group was tasked with solving was the problem of counting the number of isomers of a class of chemical compounds known as the alkanes, that are given by the chemical formula

\[ C_nH_{2n+2} \]

This is an interesting problem with an obvious application in chemistry.
Counting the monosubstituted alkanes

In order to figure out the more advanced problem we first considered rooted trees with each vertex having degree at most 3. For further simplification we considered planar trees, i.e. symmetric trees were counted as distinct. Let \( a_n \) denote the number of such trees with exactly \( n \) vertices. We define the generating polynomial (power series) as:

\[
f(x) = a_0 + a_1x + a_2x^2 + \ldots
\]

Then notice that \( a_n = \sum_{i+j=n-1} a_ia_j \). Which yields the constraint

\[
f(x) = xf(x)^2 + 1
\]

Solving this yields:

\[
f(x) = \frac{1 - \sqrt{1 - 4x}}{2x}
\]
Rearranging and expanding using Newton’s generalized binomial theorem gives:

\[ f(x) = \sum_{k \geq 0} \frac{1}{k + 1} \binom{2k}{k} x^k \]

It is worth pointing out that \( \frac{1}{n+1} \binom{2n}{n} \) the coefficients of \( f(x) \) are the celebrated Catalan numbers.
Now we consider the rooted planar trees with each vertex having degree at most 4. Again, let $a_n$ denote the number of such trees with exactly $n$ vertices. We define the generating polynomial (power series) as:

$$f(x) = a_0 + a_1x + a_2x^2 + ...$$

Then notice that $a_n = \sum_{i+j+k=n-1} a_ia_ja_k$. Which yields the constraint

$$f(x) = xf(x)^3 + 1$$

We can use the recurrence relation to generate the coefficients.
Up to now we have considered only the planar trees. Now we must eliminate the symmetric trees. It was Pólya who popularized counting ordered colourings on a set using group theory to account for symmetry. Applying Pólya’s theory of enumeration yields:

\[ f(x) = 1 + xZ_{S_2}(f(x)) \]

where

\[ Z_{S_2}(f(x)) = \frac{1}{2}(f(x)^2 + f(x^2)) \]

Which gives the recurrence relation:

\[ a_n = \frac{1}{2} \left( \sum_{i+j=n-1} a_ia_j + \sum_{2i=n-1} a_i \right) \]
Now we count the distinct rooted ternary trees (accounting for symmetry). Similarly to before we can derive:

\[ f(x) = 1 + xZ_{S_3}(f(x)) \]

where:

\[ Z_{S_3}(f(x)) = \frac{1}{6}(f(x)^3 + 3f(x)f(x^2) + 2f(x^3)) \]

Then we have the recurrence relation:

\[ a_{n+1} = \frac{1}{6} \left( \sum_{i+j+k=n} a_ia_ja_k + 3 \sum_{i+2j=n} a_ia_j + 2 \sum_{3i=n} a_i \right) \]
Plotting $n = 0, 1, 2, \ldots, 300$ against the log of our results we get the following:
Log plot of the number of structural isomers for $C_nH_{2n+1}X$
Log plot of the number of structural isomers for $C_nH_{2n+1}X$

- Log of sequence
- Fitted line
And finally we tackled the main problem: counting the number of ternary unrooted trees (accounting for symmetry). Using the center method define:

\[ f_0 = 1, \quad f_k = 1 + Z_{S_3}(f) \text{ for } k \geq 1 \]

\[ F_k = x(Z_{S_4}(f_k - f_{k-1}) + Z_{S_3}(f_k - f_{k-1})(f_{k-1}) + Z_{S_2}(f_k - f_{k-1})Z_{S_2}(f_{k-1})) \]

\[ G_k = Z_{S_2}(f_{k+1} - f_k) \]

And then the actual generating polynomial is:

\[ T(x) = 1 + x + \sum_{k \geq 1} (G_k(x) + F_k(x)) \]
Or equivalently using the centroid method, we let $F_n$ denote the number of isomers of $C_nH_{2n+1}X$ and

$$B(x) = 1 + \sum_{0 \leq k \leq \frac{n-1}{2}} F_k x^k$$

Then we have the generating polynomial:

$$T(x) = \frac{1}{24} \left( B(x^4) + 6B(x^2)B(x)^2 + 8B(x)B(x^3) + 3B(x^2)^2 + B(x^4) \right)$$

And for bicentroid trees:

$$G(x) = \frac{1}{2} \left( B(x^2) + B(x)^2 \right) - B(x)$$
Here are the results for the number of isomers of $C_nH_{2n+2}$ for $n = 0, 1, ..., 12$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>18</td>
<td>35</td>
<td>75</td>
<td>159</td>
<td>355</td>
<td>802</td>
</tr>
</tbody>
</table>
P. Rohani, O. Miramontes, M. P. Hassell
Quasiperiodicity and chaos in population models.