MORE EXAMPLES OF MARKING COMBSTRUCT GRAMMARS

Marni Mishna July 1997

This worksheet continues to explore the possibilities with combstruct[mark]. There is an introductory worksheet that explains the idea of marking and contains some basic examples which are built upon here.

```
> with(combstruct);
```

[*allstructs*, *count*, *draw*, *finished*, *gfeqns*, *gfseries*, *gfsolve*, *iterstructs*, *mark*, *moment*, *nextstruct*, *prog_gfeqns*, *prog_gfseries*, *prog_gfsolve*, *variance*]

Lowest Common Ancestor in Binary Trees

Given any two nodes in a tree there is a shortest distance between them through a nearest common ancestor. Determining the average distance between two points (similarly, the average distance to the nearest common ancestor) is a fundamental problem in combinatorics with applications in combinatorial optimization, compiling and parrallelism. We can determine this value for binary trees by considering a method similar to determining pathlength.

Average Distance Between external Nodes

Let us ease into the topic with a specialisation of the problem: nearest common ancestor among external nodes.
First consider the case where the nearest common ancestor is the root. We can graft the solution into the leaf of a larger tree to represent the more general case.
We wish to describe a bijection that counts all distances between the left children of the root and the right. We will do this by first considering a tree as stretched out, thus the product of a sequence of trees, and a node (the root). Based on our restriction, the two end trees of the sequence will be trees consisting of only a leaf. The root can occur anywhere along the sequence and to mark its place we use a sequence with one tree marked. We will say that the root appears after this node. (We will add separately the case where it occurs before). We will use a bivariate generating function to do the counting of the length of the sequence (and hence the distance between the two leaves).

```
> Bintree:={T=Union(L, Prod(N, T, T)), L=Epsilon, N=Atom,
  Chain=Prod(A, Sequence(B), A), A=Prod(len,L), B= Prod(len,
  N,Union(Prod(Epsilon,T),Prod(T,Epsilon))), len=Epsilon}:
```

We mark the point in the sequence where the lowest common ancestor goes:

```
> B1:= mark(Bintree, unlabelled, [B,m1], 1):
```

We mark the leaves of the trees for grafting purposes:

```
> B2:= mark(B1, unlabelled, [L[0],m2], 1):
```

Next, we create the bijection. It is a product of the tree to graft into (with a leaf marked telling us where), the ancestor node, and the sequence with either a spot marked for the ancestor node or none (indicating it is at the beginning of the sequence) :

```
> BCount:= B2 union
  {AncTree=Prod(T[0][1], N[0][0], Union(Chain[1][0],
  Chain[0][0]))}:
```

If we count the number of AncTrees of size *n* it will be the number of ways to choose two external nodes in all binary trees of size *n.*

```
> count([AncTree, BCount, unlabelled], size =3);
```

$$30$$

To count the distance we can sum up all of the occurrences of "*len*" in all of the AncTrees. We do this by determining the bivariate generating function.

```
> BGF:=gfsolve(BCount, unlabelled, x, [[u, len[0][0]]]):
> AT:=subs(BGF,AncTree(x,u));
```

$$AT := \frac{u^2\,x}{\sqrt{1-4\,x}\,(1-2\,u\,(1-\sqrt{1-4\,x})-4\,u^2\,x+2\,u^2\,(1-\sqrt{1-4\,x}))}$$

To give each *u* the corresponding weight, we differentiate and set *u*=1.

```
> CountLength:=eval(subs(u=1, diff(AT, u)));
```

$$CountLength := 2\,\frac{x}{(1-4\,x)^{3/2}} - \frac{x\,(2-2\,\sqrt{1-4\,x}-8\,x)}{(1-4\,x)^{5/2}}$$

This is the generating function for the sum of all shortest paths between leaves. We can find the coefficients with the series expansion, and in fact we can determine the closed form for the coefficients, using gfun.

```
> series(CountLength, x, 10);
```

$$2\,x+16\,x^2+96\,x^3+512\,x^4+2560\,x^5+12288\,x^6+57344\,x^7+262144\,x^8+1179648\,x^9+\mathrm{O}(x^{10})$$

```
> with(gfun):
  T1:=
  expand(rsolve(diffeqtorec(holexprtodiffeq(CountLength,
  y(x)), y(x), u(n)), u(n)));
```

$$T1 := \frac{1}{2}\,4^n\,n$$

To find the average sum per tree we divide by the number of trees of size *n*. This is represented in the grammar as T[0][0]. To find the average distance we also need to divide by the number of pairs of leaves. A tree with *n* nodes has *n*+1 leaves and thus this is $(n+1)\ \mathit{Choose}\ 2 = \frac{(n+1)\,n}{2}$.

```
> Total:=rsolve(diffeqtorec(holexprtodiffeq(subs(BGF,T[0][0]
  (x,u)), y(x)), y(x), u(n)), u(n));
```

$$Total := \frac{4^n\,\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+2)\,\sqrt{\pi}}$$

```
> simplify(T1/Total/((n+1)*n/2),GAMMA);
```

$$\frac{\sqrt{\pi}\,\Gamma(n+1)}{\Gamma\left(n+\frac{1}{2}\right)}$$

Here are the values for a few small trees:

```
> seq(evalf(T1/Total/((n+1)*n/2)), n=1..10);
```

2., 2.666666667, 3.200000000, 3.657142857, 4.063492063, 4.432900433, 4.773892774, 5.092152292, 5.391690662, 5.675463855

The distance to the nearest common ancestor is half of this value since the trees have a symmetric construction.

Asymptotically we have:

```
> map(combine, asympt(T1/Total/((n+1)*n/2), n), exp);
```

$$\frac{\sqrt{\pi}}{\sqrt{\frac{1}{n}}} + \frac{1}{8}\sqrt{\pi}\sqrt{\frac{1}{n}} + \frac{1}{128}\sqrt{\pi}\left(\frac{1}{n}\right)^{3/2} - \frac{5}{1024}\sqrt{\pi}\left(\frac{1}{n}\right)^{5/2} + O\left(\left(\frac{1}{n}\right)^{7/2}\right)$$

Average Distance Between any two Nodes

To extend the problem to nearest common ancestor between any two nodes we need to allow the end bits of our sequence to be more than just leaves, but in fact trees. This is attained with a simple substitution. We allow *A* to be a tree. This omits the case when one of the nodes is the ancestor. This case is omitted since it is not possible for a node to be both an ancestor of a second node and a leaf of the same tree, and this is precisely what is needed. To account for this, we add the possibility of a second chain, much like the first except it is terminated on the top end by a node (not a tree). To avoid over counting by one we do not add a tag on the end tree.

```
> Bintree:={T=Union(L, ST), ST= Prod(N, T, T), L=Epsilon,
  N=Atom,
  Chain=Prod(A, Sequence(B), A), A=Prod(len,T), B= Prod(len,
  N,Union(Prod(Epsilon,T),Prod(T,Epsilon))), Chain2=Prod(T,
  Sequence(B, card>0)), len=Epsilon}:
```

We mark the point in the sequence where the node goes:

```
> B1:= mark(Bintree, unlabelled, [B,m1], 1):
```

We mark the leaves of the trees for grafting purposes:

```
> B2:= mark(B1, unlabelled, [L[0],m2], 1):
```

Next, we create the bijection. It is a product of the tree to graft into (with a leaf marked telling us where), the ancestor node, the sequence with either a spot marked for the ancestor or none (indicating it is at the beginning of the sequence). Or it is the chain that represents the case when one of the nodes is the ancestor.

```
> BCount:= B2 union
  {AncTree=Union(Prod(T[0][1], N[0][0], Union(Chain[1][0],
  Chain[0][0])),PL),
   PL=Prod(T[0][1], Chain2[0][0])}:
```

If we count the number of AncTrees of size *n* it will be the number of ways to choose two nodes or leaves from a binary tree of size *n*.

```
> count([AncTree, BCount, unlabelled], size = 2);
```

20

As before, to count the distance we can sum up all of the occurrences of "*len*" in all of the AncTrees. We can solve the bivariate generating function.

```
> BGF:=gfsolve(BCount, unlabelled, x, [[u, len[0][0]]]):
```

```
> AT:=subs(BGF,AncTree(x,u));
```

$$AT := u\left(7\,u^2\,(1-\sqrt{1-4\,x}) - 7\,u + 2 - 4\,u\,(1-\sqrt{1-4\,x}) - 8\,u^2\,x + 6\,u^2 - 3\,\frac{u^2\,(1-\sqrt{1-4\,x})}{x} + \frac{7}{2}\,\frac{u\,(1-\sqrt{1-4\,x})}{x} - \frac{1-\sqrt{1-4\,x}}{x}\right) \Big/ \left(\sqrt{1-4\,x}\,(1-2\,u\,(1-\sqrt{1-4\,x}) - 4\,u^2\,x + 2\,u^2\,(1-\sqrt{1-4\,x}))\,(-1+u\,(1-\sqrt{1-4\,x}))\right)$$

```
> CountLength:=eval(subs(u=1, diff(AT, u)));
```

$$CountLength := -\frac{4 - 3\,\sqrt{1-4\,x} - 8\,x - \frac{1}{2}\,\frac{1-\sqrt{1-4\,x}}{x}}{(1-4\,x)^2} - \frac{15 - 10\,\sqrt{1-4\,x} - 16\,x - \frac{5}{2}\,\frac{1-\sqrt{1-4\,x}}{x}}{(1-4\,x)^2} + \frac{\left(4 - 3\,\sqrt{1-4\,x} - 8\,x - \frac{1}{2}\,\frac{1-\sqrt{1-4\,x}}{x}\right)(2 - 2\,\sqrt{1-4\,x} - 8\,x)}{(1-4\,x)^3} - \frac{\left(4 - 3\,\sqrt{1-4\,x} - 8\,x - \frac{1}{2}\,\frac{1-\sqrt{1-4\,x}}{x}\right)(1-\sqrt{1-4\,x})}{(1-4\,x)^{5/2}}$$

```
> series(CountLength, x, 10);
```

$$4\,x + 36\,x^2 + 232\,x^3 + 1300\,x^4 + 6744\,x^5 + 33320\,x^6 + 159184\,x^7 + 742068\,x^8 + \mathrm{O}(x^9)$$

Again, we can find a closed form for the coefficients using gfun:

```
> T1:=rsolve(diffeqtorec(holexprtodiffeq(CountLength, y(x)),
  y(x), u(n)), u(n));
```

$$T1 := -4\,\frac{4^n \left(\frac{1}{2}\right)^n 2^n\,\Gamma\!\left(\frac{3}{2}+n\right)}{\Gamma(n+1)\,\sqrt{\pi}} + 2\,4^n + 2\,4^n\,n$$

To find the the average, we again need to divide by the number of trees of size n and the number of ways of choosing two nodes. In this case, since we are including internal nodes there are a total of *2n+1* nodes, and thus there are *n(2n+1)* ways of choosing two.

```
> Total:=rsolve(diffeqtorec(holexprtodiffeq(subs(BGF,T[0][0]
  (x,u)), y(x)), y(x), u(n)), u(n)):
```

```
> AVG:= simplify(T1/Total/n/(2*n+1), symbolic);
```

$$AVG := \frac{(n+1)\left(-2\,\Gamma\left(\frac{3}{2}+n\right)+\sqrt{\pi}\,\Gamma(n+2)\right)}{n\,\Gamma\left(\frac{3}{2}+n\right)}$$

```
> seq(evalf(AVG), n=1..10);
```

1.333333333, 1.800000000, 2.209523810, 2.579365079, 2.919480519, 3.236208236, 3.533888334, 3.815651995, 4.083848728, 4.340296442

Notice that these values are smaller than when leaves alone were considered. We can also examine the asymptotic behaviour:

```
> map( simplify, asympt(AVG, n), symbolic);
```

$$\sqrt{\pi}\,\sqrt{n}-2+\frac{13}{8}\frac{\sqrt{\pi}}{\sqrt{n}}-\frac{2}{n}+\frac{57}{128}\frac{\sqrt{\pi}}{n^{3/2}}-\frac{89}{1024}\frac{\sqrt{\pi}}{n^{5/2}}+O\left(\frac{1}{n^{7/2}}\right)$$

Non-Crossing Configurations

To better understand the capabilities for determining mean and variance, we now consider the enumeration of planar configurations defined on vertices of a convex *n*-gon. We will look at six basic non-crossing configurations: trees and forests, graphs and connected graphs, dissections and partitions. We will use marking directly and indirectly via the commands combstruct[mean] and combstruct[variance] to determine properties of these figures. The simpler constructions offer more information, however we can determine specific moments and variances for sub-objects when the bivariate generating function is not available in closed form.

Trees and Forests

Trees

First we consider non-crossing trees and non-crossing forests. A **tree** is a sequence of butterflies attached to a root, where a **butterfly** is an ordered pair of trees whose roots have been merged into a single node. This merge step may duplicate the root, so a butterfly is expressed as the product of two forests and a root, where a **forest** is a sequence of butterflies. We are interested in the average number of leaves in such a tree.

```
> nctree:={T=Union(L, Prod(Z, Sequence(B, card>0))),
  F=Sequence(B, card>0), B=Union(Prod(F,Z,F), Prod(F, Z),
  Prod(Z, F), L), L=Atom}:
> marked_trees:= mark(nctree, labelled, [L, m], 2):
```

We can use the moment command to give us specific moments. The first moment is the average.

```
> evalf(moment([T, nctree, labelled], size =25, L,
  1)/25);
```

.4416901408

Notice that this value is equal to the {*number trees of size* n *with one leaf marked*}/{*number of trees of size* n}.

```
> evalf(count([T[1], marked_trees, labelled]
  ,size=25)/count([T[0], marked_trees,labelled],
  size=25))/25;
```

$$.4416901408$$

We can determine the general form of the coefficients of the generating functions of the marked productions and hence determine a closed form for the mean percentage of leaves in a tree of size *n*.

```
> TT:=gfsolve(marked_trees, labelled, z):
```

```
> T1:=rsolve(diffeqtorec(holexprtodiffeq(subs(TT,
  T[1](z)), y(z)), y(z), u(n)) minus {u(1) = 1, u(0)=0},
  u(n));
```

$$T1 := \frac{1}{486}\,\frac{27^n\sqrt{3}\,\Gamma\left(n-\frac{2}{3}\right)\Gamma\left(n-\frac{4}{3}\right)}{\pi\,\Gamma(2\,n-2)}$$

```
> T0:=rsolve(diffeqtorec(holexprtodiffeq(subs(TT,
  T[0](z)), y(z)), y(z), u(n)) minus{u(0)=0}, u(n));
```

$$T0 := \frac{1}{54}\,\frac{27^n\sqrt{3}\,\Gamma\left(n-\frac{1}{3}\right)\Gamma\left(n-\frac{2}{3}\right)}{\pi\,\Gamma(2\,n)}$$

```
> T2:=rsolve(diffeqtorec(holexprtodiffeq(subs(TT,
  T[2](z)), y(z)), y(z), u(n)) minus{u(2)=0, u(1)=0,
  u(0)=0}, u(n));
```

$$T2 := \frac{1}{8748}\,\frac{27^n\sqrt{3}\,\Gamma\left(n-\frac{4}{3}\right)\Gamma\left(n-\frac{5}{3}\right)}{\pi\,\Gamma(2\,n-4)}$$

The mean number of leaves in the tree is determined:

```
> simplify(T1/T0/n, GAMMA);
```

$$\frac{2}{3}\,\frac{(n-1)\,(2\,n-1)}{(3\,n-4)\,n}$$

```
> asympt(", n);
```

$$\frac{4}{9}-\frac{2}{27}\frac{1}{n}+\frac{10}{81}\frac{1}{n^2}+\frac{40}{243}\frac{1}{n^3}+\frac{160}{729}\frac{1}{n^4}+\frac{640}{2187}\frac{1}{n^5}+O\left(\frac{1}{n^6}\right)$$

Notice that 4/9 is near the value we had determined for the value at *n=25*.

We can also determine a closed form for the variance:

```
> evalf(variance([T, nctree, labelled], size =25, L)/25);
```

$$.1117407922$$

```
> var:=simplify(((2*T2+T1)/T0-(T1/T0)^2)/n, GAMMA);
```

$$var := \frac{2}{9}\,\frac{(n-1)\,(2\,n-1)\,(7\,n^2-25\,n+22)}{(3\,n-5)\,(3\,n-4)^2\,n}$$

```
> asympt(var, n);
```

$$\frac{28}{243}-\frac{62}{729}\frac{1}{n}-\frac{106}{2187}\frac{1}{n^2}-\frac{580}{6561}\frac{1}{n^3}+O\left(\frac{1}{n^4}\right)$$

The asymptotic forms tell us that both the mean and the variance tend to a constant.

```
> evalf(28/243);
```

.1152263374

This is similar to the value found at $n=25$ as well.

Forests

A non-crossing forest is a collection of non-crossing trees. We can obtain a forest by substituting in for each vertex of a tree a vertex and a forest, (possibly empty). We are interested in the average number of components in a forest on n nodes. Thus, we must distinguish the occurrence of a new, non-empty forest, marked by the non-terminal *comp*.

```
> fo:={B=Prod(Sequence(B),V,Sequence(B)),
       V=Prod(Z, F),
       F=Union(Epsilon, comp), comp=Prod(V,
  Sequence(B))}:
```

First, let us determine the variance and moments for a large class of forests:

```
> evalf(moment([F, fo, labelled], size =30, comp, 1))/30;
```

.2141335590

```
> evalf(variance([F, fo, labelled], size =30, comp))/30;
```

.1429128809

We have less success determining the general form with marking in this instance:

```
> mf:=mark(fo, labelled, [comp, m], 1):
> FF:=gfsolve(mf, labelled, z):
> F1:=rsolve(diffeqtorec(holexprtodiffeq(subs(FF,
  V[1](z)), y(z)), y(z), u(n)), u(n));
Error, (in LREtools[polysols]) Only linear equations are handled
```

However, we do have the generating functions at our disposal.

Connected and General Graphs

Connected graphs

We now alleviate the constraint of tree and consider more general graphs. We build the graphs in a similar way, with a few added precautions to avoid double counting of nodes. Consider the set of children $\{v_i, v_{1+i},..,v_j\}$ attached to the root v_1. This set divides the graph into several parts. We have the graphs built on $\{v_2, ..., v_i\}$ and $\{v_j, .., v_{n-1}\}$ which are connected themselves by hypothesis, and between any two adjacent children v_k and v_{k+1} there either one connected graph, or two neighbouring connected graphs.

Any graph built from a sequence of successive vertices of the circle is a system of arches since the edges of the graphs, or arcs, are not allowed to cross. The endpoints of these arches are shared by the graphs built to the left and to the right of a given v_k, so that we shall say that the size of an arch built on n points has size n-1. We are interested here in **Elementary Arches**, that is arches that always contain an arch between the first and last points. General arches are easily obtained by sequencing *EAs*.

From this, we have the following combstruct specification. In particular, the 5 arguments of a *congraph* entity are as follows:

1. First Z: the root of the graph
2. Second Z: the leftmost point of the first *EA*
3 and 4. the two *EA*s built on $\{v_2, ..., v_i\}$ and $\{v_j, .., v_{n-1}\}$
5. The sequence of *EA*s found between two consecutive children of v_1. The first term corresponds to one *EA*,
and the second one to two *EA*s. For the latter case, the Z in the Prod stands for the lefmost point of the second *EA*.

The component of interest here is edges. We are interested in the average number of edges in a connected graph. Every arch is an edge, plus the original edges to the children. Thus we distinguish these as *EA* and *conseq*.

```
> ar:={EA = Union(Sequence(EA, card >= 2),
                  Prod(Z, Sequence(EA), Sequence(EA))
                 ),
       congraph = Prod(Z,Z,Sequence(EA),
  Sequence(EA),Sequence(conseq)
                 ),
       conseq = Union(Sequence(EA,card>=1),
  Prod(Z,Sequence(EA),Sequence(EA)))}:
> marked_edges:= mark(ar, labelled, [[conseq, EA], m],
  2):
> ME:=gfsolve(marked_edges,labelled,z):
> simplify(subs(ME, congraph[1](z)),symbolic);
```

$$-z\,(2\,z+18\,z^2-5\,\mathrm{RootOf}(\_Z^3\,z+(1+3\,z)\,\_Z^2+(-1+3\,z)\,\_Z+z)\,z$$
$$-2\,\mathrm{RootOf}(\_Z^3\,z+(1+3\,z)\,\_Z^2+(-1+3\,z)\,\_Z+z)$$
$$+126\,z^2\,\mathrm{RootOf}(\_Z^3\,z+(1+3\,z)\,\_Z^2+(-1+3\,z)\,\_Z+z)$$
$$-\mathrm{RootOf}(\_Z^3\,z+(1+3\,z)\,\_Z^2+(-1+3\,z)\,\_Z+z)^2\,z)\,/\,(108\,z^2-1)$$

```
>
> P0:=rsolve(diffeqtorec(holexprtodiffeq(subs(ME,
  congraph[0](z)), y(z)), y(z),u(n)), u(n));
```

$$P0 := \mathrm{rsolve}(\{\,\mathrm{u}(1)=0,\ \mathrm{u}(0)=0,\ (432+1944\,n^2-1944\,n)\,\mathrm{u}(n)$$
$$+(-168-108\,n^2-216\,n)\,\mathrm{u}(n+1)+(-18\,n^2-36\,n)\,\mathrm{u}(n+2)$$
$$+(n^2+5\,n+6)\,\mathrm{u}(n+3),\ \mathrm{u}(3)=4,\ \mathrm{u}(4)=23,\ \mathrm{u}(2)=1\,\},\ \mathrm{u}(n))$$

We also have moments for any size class available to us:

```
> evalf(variance([congraph, ar, labelled],size =30,
  [conseq, EA])/30);
```

$$.2305481447$$

```
> evalf(moment([congraph, ar, labelled],size =50,
  [conseq, EA],1)/50);
```

$$1.305950905$$

General Graphs

We create general graphs from connected graphs the same way we obtained forests from trees. A general graph is obtained from a connected one by the substitution Z -> Prod(Z, G). So that we just have to rewrite the previous grammar by adding a new

symbol which makes this substitution.
However the decomposition of a connected graph misses a configuration for general graphs: the one where vertex v_1 has no child. We are interested in two parameters in this instance: the number of edges and the number of components.

```
> br:={EA = Union(Sequence(EA, card >= 2),
                  Prod(V, Sequence(EA), Sequence(EA))
                 ),
       V=Union(Z,Comp), Comp=Prod(Z, G),
       G=Union(Z,
               Comp,
               Prod(V,V,Sequence(EA),
  Sequence(EA),Sequence(F))),
       F=Union(Sequence(EA,card>=1),
  Prod(V,Sequence(EA),Sequence(EA)))
      }:
> marked_edges:= mark(br, labelled, [G, m], 2):
```

We attempt to solve the generating function:

```
> GE:= gfsolve(marked_edges, labelled, z):
> subs(GE, G[1](z));
```

$$-\Bigg(4980734\,z+1424400540\,z^4-61866114\,z^2$$
$$+52428544\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^9}{1+3z}$$
$$+3239936\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^{13}}{1+3z}$$
$$+36674304\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^8}{1+3z}$$
$$+26034176\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^{11}}{1+3z}$$
$$+\frac{66322501}{2}\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^2}{1+3z}$$
$$-121109347\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^3}{1+3z}$$
$$+43130368\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^{10}}{1+3z}$$
$$+450560\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^{14}}{1+3z}$$
$$+6201482770\,\frac{(3+2z-\sqrt{4z^2-12z+1})\,z^5}{1+3z}$$

$$- 689149615 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^4}{1 + 3z}$$

$$- \frac{5111807}{2} \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z}{1 + 3z}$$

$$- 14190069348 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^6}{1 + 3z}$$

$$+ 9261438168 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^7}{1 + 3z}$$

$$+ 11382784 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^{12}}{1 + 3z} + 197128100\, z^3 - 901120\, z^{13}$$

$$- 55984128\, z^{10} - 23216128\, z^{11} - 6479872\, z^{12} - 10987396280\, z^5$$

$$+ 22448267792\, z^6 - 12620470560\, z^7 - 161705472\, z^8 - 103967744\, z^9$$

$$+ 65536 \frac{3 + 2z - \sqrt{4z^2 - 12z + 1}}{1 + 3z} - 131072 \Bigg) z \Bigg/ \Bigg(\Bigg(\Bigg($$

$$\frac{63}{2} \frac{3 + 2z - \sqrt{4z^2 - 12z + 1}}{1 + 3z} - 15 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z}{1 + 3z}$$

$$- 18 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^2}{1 + 3z} + 4 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^3}{1 + 3z}$$

$$+ 24 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^4}{1 + 3z} - 62 - 12z - 24z^2 - 48z^3 - 32z^4$$

$$+ 16 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^5}{1 + 3z} \Bigg) (1 + 3z) \Bigg(\frac{15}{2} \frac{3 + 2z - \sqrt{4z^2 - 12z + 1}}{1 + 3z}$$

$$+ 5 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z}{1 + 3z} + 6 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^2}{1 + 3z} - 14 - 12z$$

$$- 8z^2 + 4 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^3}{1 + 3z} \Bigg) \Bigg($$

$$78012 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^7}{1 + 3z} - 104976\, z^7 + 403848\, z^6$$

$$- 278274 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^6}{1 + 3z}$$

$$+ 119743 \frac{(3 + 2z - \sqrt{4z^2 - 12z + 1})\, z^5}{1 + 3z} - 242436\, z^5$$

$$+\frac{179729}{2}\frac{(3+2z-\sqrt{4z^2-12z+1})z^4}{1+3z}-138650z^4+36850z^3$$

$$-\frac{33669}{2}\frac{(3+2z-\sqrt{4z^2-12z+1})z^3}{1+3z}-\frac{9245}{2}\frac{(3+2z-\sqrt{4z^2-12z+1})z^2}{1+3z}$$

$$+7922z^2-1598z+\frac{1663}{2}\frac{(3+2z-\sqrt{4z^2-12z+1})z}{1+3z}+64$$

$$\left.\left.-32\frac{3+2z-\sqrt{4z^2-12z+1}}{1+3z}\right)\right)$$

Unfortunately, again rsolve is unable to determine an expression for the coefficients. However, we can solve for specific instances:

```
> evalf(variance([G, br, labelled],size =30, [F,
  EA])/30);
  evalf(moment([G, br, labelled],size =30, [F,
  EA],1)/30);
```

$$.5060548662$$

$$1.078335530$$

Next, we handle components. Each occurrence of *comp* marks a new component.

```
> marked_comps:= mark(br, labelled, [Comp,m], 2):
> GC:= gfsolve(marked_comps, labelled, z):
```

```
> subs(GC, G[1](z));
```

$$\left(24576\frac{(3+2z-\sqrt{4z^2-12z+1})z^8}{1+3z}-49152z^7\right.$$

$$+225280\frac{(3+2z-\sqrt{4z^2-12z+1})z^7}{1+3z}$$

$$-100902080\frac{(3+2z-\sqrt{4z^2-12z+1})z^6}{1+3z}+135598336z^6$$

$$+199862896\frac{(3+2z-\sqrt{4z^2-12z+1})z^5}{1+3z}-308403776z^5+189557280z^4$$

$$-109317736\frac{(3+2z-\sqrt{4z^2-12z+1})z^4}{1+3z}-37693664z^3$$

$$+19905992\frac{(3+2z-\sqrt{4z^2-12z+1})z^3}{1+3z}$$

$$-1049424\frac{(3+2z-\sqrt{4z^2-12z+1})z^2}{1+3z}+2100096z^2$$

$$\left.-29\frac{(3+2z-\sqrt{4z^2-12z+1})z}{1+3z}+108z-\frac{1}{2}\frac{3+2z-\sqrt{4z^2-12z+1}}{1+3z}+2\right)z^2$$

$$
\Big/ \left(\left(949392 \frac{(3+2z-\sqrt{4z^2-12z+1})z^7}{1+3z} - 1259712 z^7 \right.\right.
$$

$$
- 2018616 \frac{(3+2z-\sqrt{4z^2-12z+1})z^6}{1+3z} + 3140064 z^6
$$

$$
- 218032 \frac{(3+2z-\sqrt{4z^2-12z+1})z^5}{1+3z} - 56832 z^5
$$

$$
+ 518120 \frac{(3+2z-\sqrt{4z^2-12z+1})z^4}{1+3z} - 968224 z^4
$$

$$
+ 60279 \frac{(3+2z-\sqrt{4z^2-12z+1})z^3}{1+3z} - 83044 z^3
$$

$$
- \frac{33557}{2} \frac{(3+2z-\sqrt{4z^2-12z+1})z^2}{1+3z} + 34042 z^2
$$

$$
\left. - 20 \frac{(3+2z-\sqrt{4z^2-12z+1})z}{1+3z} + 72 z - \frac{1}{2} \frac{3+2z-\sqrt{4z^2-12z+1}}{1+3z} + 2 \right) \Bigg(
$$

$$
\frac{15}{2} \frac{3+2z-\sqrt{4z^2-12z+1}}{1+3z} + 5 \frac{(3+2z-\sqrt{4z^2-12z+1})z}{1+3z}
$$

$$
+ 6 \frac{(3+2z-\sqrt{4z^2-12z+1})z^2}{1+3z} - 14 - 12z - 8z^2
$$

$$
\left.\left. + 4 \frac{(3+2z-\sqrt{4z^2-12z+1})z^3}{1+3z} \right)\right)
$$

Again, the closed form for coefficients is beyond the capabilities of rsolve. So, we look at specific examples:

```
> evalf(variance([G, br, labelled],size =30, Comp)/30);
  evalf(moment([G, br, labelled],size =30, Comp,1)/30);
```

$$.1001034427$$

$$.1130857137$$

Dissections and Partitions

Dissections

A dissection of a convex polygon P_n={v_1, ..., v_n} is a partition of the polygon into polygonal regions by means of non-intersecting diagonals. Our representation is the same as non-crossing graphs whose connected components are points (polygons on one point) , edges (polygons on two points) and cycles (all polygons). We are curious to determine the average number of parts in a dissection.

```
> dissG:={Di=Union(Z, P),  P=Sequence(Di, card >= 2)}:
> marked_dp:= mark(dissG, labelled, [[Z, P], m], 2):
  MD:=gfsolve(marked_dp, labelled, z);
```

$$MD := FAIL$$

A more careful look at this failure shows that it is actually due to a weakness in series. This can be overcome by increasing Order:

```
> Order:=20:
> MD:=gfsolve(marked_dp, labelled, z):
```

Unfortunately the recurrence set up by gfun is unsolveable using `rsolve`, so we cannot attain a closed form for the coefficients. However, we do have the generating function:

```
> simplify(subs(MD, Di[1](z)), power);
```

$$\frac{1}{4}\left(-1+2z+\sqrt{1-6z+z^2}+4z^2+z\sqrt{1-6z+z^2}-6z^3+3z^2\sqrt{1-6z+z^2}+z^4-z^3\sqrt{1-6z+z^2}\right)\Big/\left(1-5z-\sqrt{1-6z+z^2}-5z^2+2z\sqrt{1-6z+z^2}+z^3-z^2\sqrt{1-6z+z^2}\right)$$

```
> evalf(variance([Di, dissG, labelled], size =75, [P,
  Z])/75);
```

$$.1730906544$$

```
> evalf(moment([Di, dissG, labelled], size =75, P,
  1)/75);
```

$$.7007320871$$

Partitions

A non crossing partition of size n is a partition of $[n]=\{1, 2, ..., n\}$ such that if $a < b < c < d$ and a block contains a and c, then no block contains b and d. Thus, if we consider each partition block as a polygon, partitions are the "forest" version of dissections. Given a partition block, possibly empty, one gets a bigger partition by substituting to each vertex the product between
Z and another partition. We are interested in the average number of partition blocks in a partition.

```
> partG:={P=Sequence(V, card>0), V = Prod(Z,Union(P,
  Epsilon))}:
> marked_parts:= mark(partG, labelled, [P, m], 2):
  MP:=gfsolve(marked_parts, labelled, z):
```

We try to solve for the coefficients explicitly:

```
> P1:=rsolve(diffeqtorec(holexprtodiffeq(subs(MP,
  P[1](z)), y(z)), y(z),u(n)) minus{u(0)=0}, u(n));
  P0:=rsolve(diffeqtorec(holexprtodiffeq(subs(MP,
  P[0](z)), y(z)), y(z),u(n)) minus{u(0)=0}, u(n));
  P2:=rsolve(diffeqtorec(holexprtodiffeq(subs(MP,
  P[2](z)), y(z)), y(z), u(n)) minus{u(2)=0, u(1)=0,
  u(0)=0}, u(n));
```

$$P1 := \frac{1}{2}\,\frac{4^n\,\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+1)\sqrt{\pi}}$$

$$P0 := \frac{4^n \, \Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+2)\sqrt{\pi}}$$

$$P2 := \frac{1}{8}\frac{4^n \, \Gamma\left(n-\frac{1}{2}\right)}{\Gamma(n-1)\sqrt{\pi}}$$

```
> simplify(P1/P0/n, GAMMA);
  asympt(", n);
```

The mean tends to the constant 1/2.

$$\frac{1}{2}\frac{n+1}{n}$$

$$\frac{1}{2}+\frac{1}{2}\frac{1}{n}$$

```
> var:=simplify(((2*P2+P1)/P0-(P1/P0)^2)/n, GAMMA);
  asympt(var, n);
```

The variance tends to the constant 1/8.

$$var := \frac{1}{4}\frac{(n-1)(n+1)}{n(2n-1)}$$

$$\frac{1}{8}+\frac{1}{16}\frac{1}{n}-\frac{3}{32}\frac{1}{n^2}-\frac{3}{64}\frac{1}{n^3}-\frac{3}{128}\frac{1}{n^4}-\frac{3}{256}\frac{1}{n^5}-\frac{3}{512}\frac{1}{n^6}-\frac{3}{1024}\frac{1}{n^7}-\frac{3}{2048}\frac{1}{n^8}$$
$$-\frac{3}{4096}\frac{1}{n^9}-\frac{3}{8192}\frac{1}{n^{10}}-\frac{3}{16384}\frac{1}{n^{11}}-\frac{3}{32768}\frac{1}{n^{12}}-\frac{3}{65536}\frac{1}{n^{13}}-\frac{3}{131072}\frac{1}{n^{14}}$$
$$-\frac{3}{262144}\frac{1}{n^{15}}-\frac{3}{524288}\frac{1}{n^{16}}-\frac{3}{1048576}\frac{1}{n^{17}}-\frac{3}{2097152}\frac{1}{n^{18}}-\frac{3}{4194304}\frac{1}{n^{19}}$$
$$+\mathrm{O}\left(\frac{1}{n^{20}}\right)$$

```
>
```