Counting Join Trees

MMF, version of November 3, 2008, 11:09

Problem. Let *Obj* be a set of so-called objects. Let *Tree* be the set of non-empty finite binary trees over *Obj* with *tip* as tip former and $\oplus$ as node constructor:

$$\begin{array}{lcl} tip & : & Obj \to Tree \\ \oplus & : & Tree \times Tree \to Tree \end{array}$$

We define function *tips* to yield the sequence of tips of a tree:

$$\begin{array}{lcl} tips & : & Tree \to \mathrm{seq}\, Obj \\ tips(tip(o)) & = & \langle o \rangle \\ tips(x \oplus y) & = & tips(x) \frown tips(y) \end{array}$$

We have used the notation $\langle x \rangle$ for a singleton sequence, $\frown$ for sequence concatenation, and we consider a sequence s to be total function from $0 \mathinner{..} \#s$ to the set of elements contained in the sequence, so that unary operation ran effectively is the sequence-to-set conversion: $\langle x, y, z \rangle = \{(0, x), (1, y), (2, z)\} \overset{\mathrm{ran}}{\mapsto} \{x, y, z\}$.

Define equivalence relation $\simeq$ on *Tree* by:

$$\begin{array}{lcl} t \oplus t' & \simeq_1 & t' \oplus t \\ (\simeq) & = & \text{reflexive, symmetric, transitive, and } (tip, \oplus)\text{-congruent closure of } (\simeq_1) \end{array}$$

Thus trees are equivalent iff they can be transformed into each other by repeatedly interchanging the arguments of (the topmost or some internal) $\oplus$-nodes in the tree. The equivalence class of a tree t is denoted $[t]_\simeq$ or just $[t]$:

$$[t] \quad = \quad \{t' : Tree \mid t' \simeq t\}$$

Lemma: $\#[t] = 2^n$ where $n{+}1 = \#tips(t)$. (Proof sketch: for each $\oplus$-node, rule $\simeq_1$ is applicable; there are n $\oplus$-nodes in a tree with $n{+}1$ tips.)

Now the question is: what is the number of non-equivalent ways to join $n{+}1$ different objects? The answer is given by the following theorem.

Theorem. Let O be a set of $n{+}1$ different objects, $n \geq 0$. Then:

$$\#\{t : Tree \mid \mathrm{ran}\, tips(t) = O \bullet [t]\} \quad = \quad \frac{(n{+}1)!}{2^n} Cat_n \quad = \quad \frac{n!}{2^n} \binom{2n}{n}$$

Here, Cat_n is the n-th *Catalan* number [1, 2, 3]; it denotes the number of "binary parenthezations" of a sequence s of $n{+}1$ different objects: $Cat_n = \#\{t \mid tips(t) = s\} = \frac{1}{n+1}\binom{2n}{n}$.

After having discovered (by googling on 'parentheses binary tree leaves number count') the Catalan numbers Cat_n [1], I readily thought of some claims like in the theorem. In the end, it was not hard to be convinced of the truth of the theorems's claim, by an informal argument. The proof below *closely* follows the informal argument; its elegant formulation (as a *calculation*) took me much more time than the time I needed to get the idea.

Proof. We omit the domain indications when no confusion can arise, tacitly assuming that all sequences s are Obj sequences of length $n{+}1$. Almost all steps are quite simple; only the three $\stackrel{!}{=}$-steps require some ingenuity.

$$\#\{t \mid \operatorname{ran} tips(t) = O \bullet [t]\}$$

$\stackrel{!}{=}$
- **eureka**: "instead of counting the classes, just count *all* their members and then divide by the average class size". Formally, and slightly more specific:
- law: $\#\{A, B, C \ldots\} = \frac{1}{k} \times \#\{A \cup B \cup C \ldots\}$ if $k = \#A = \#B = \#C \ldots$ and the sets A, B, C... are mutually disjoint
- equivalences classes are mutually disjoint
- above lemma: all $[t]$ have size 2^n; $\boxed{\textbf{put } k = 2^n}$

$$\tfrac{1}{k} \times \#\{t \mid \operatorname{ran} tips(t) = O \bullet t\}$$

$=$ notational convention

$$\tfrac{1}{k} \times \#\{t \mid \operatorname{ran} tips(t) = O\}$$

$=$ (*very very* simple) set calculus: $\operatorname{ran} s = X \quad \Leftrightarrow \quad s \in \{s \mid \operatorname{ran} s = X\}$

$$\tfrac{1}{k} \times \#\{t \mid tips(t) \in \{s \mid \operatorname{ran} s = O\}\}$$

$=$

$$\tfrac{1}{k} \times \#\{t \mid (\textstyle\bigvee_{s \mid \operatorname{ran} s = O} tips(t){=}s)\}$$

$=$

$$\tfrac{1}{k} \times \#(\textstyle\bigcup_{s \mid \operatorname{ran} s = O}\{t \mid tips(t) = s\})$$

$(*) =$
- law: for disjoint A, B we have $\#(A \cup B) = \#A + \#B$
- $\{t \mid tips(t) = s\}$ is disjoint from $\{t \mid tips(t) = s'\}$ for $s \neq s'$ (explained below)

$$\tfrac{1}{k} \times \textstyle\sum_{s \mid \operatorname{ran} s = O} \#\{t \mid tips(t) = s\}$$

$\stackrel{!}{=}$ **Catalan**: $\#\{t \mid tips(t) = s\}$ depends only on $\#s$, which is $n{+}1$, and equals Cat_n

$$\tfrac{1}{k} \times \textstyle\sum_{s \mid \operatorname{ran} s = O} Cat_n$$

$\stackrel{!}{=}$ **combinatorics**: there are $(n{+}1)!$ sequences each of which has O as its range

$$\tfrac{1}{k} \times (n{+}1)! \times Cat_n$$

The hint of step $(*)$ above is almost trivial; its formal proof reads as follows:

$$\{t \mid tips(t) = s\} \text{ is disjoint from } \{t \mid tips(t) = s'\}$$

$=$

$$\forall\, t, t' \mid tips(t) = s \wedge tips(t') = s' \bullet t \neq t'$$

$=$

$$\forall\, t, t' \mid t = t' \bullet \neg\,(tips(t) = s \wedge tips(t') = s')$$

$=$

$$\forall\, t \bullet \neg\,(s = tips(t) = s')$$

$\Leftarrow$ $tips$ is a function

$$s \neq s'$$

References

[1] Louis W. Shapiro and Robert A. Sulanke. Bijections for the Schröder Numbers. *Mathematics Magazine*, 73(5):369–376, 2000.

[2] R.P. Stanley. *Enumerative Combinatorics Volume II*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, UK, 1999.

[3] Wikipedia. Catalan number — wikipedia, the free encyclopedia, 2008. `http://en.wikipedia.org/w/index.php?title=Catalan_number&oldid=24456867%8`; accessed 3-Nov-2008.