Counting Trees

Marko Petkovšek*

Tomaž Pisanski†

IMFM, Department of Theoretical Computer Science

Ljubljana, Slovenia

Let t_n denote the number of unlabeled trees on n vertices. Let $t(x) = \sum_{n=1}^{\infty} t_n x^n$ be the corresponding generating function. Similarly, let T_n, h_n, and i_n denote the numbers of rooted trees, homeomorphically irreducible trees, and identity trees on n vertices, respectively. (*Homeomorphically irreducible trees* have no vertices of degree 2, and *identity trees* have trivial automorphism group.) Then their respective generating functions $T(x)$, $h(x)$, and $i(x)$ satisfy the following identities (see Harary [1], Eqns. (15.35), (15.41), (15.47 – 49), and (15.51 – 52); see also remarks following our table of i_n):

$$T(x) = x \exp \sum_{r=1}^{\infty} \frac{T(x^r)}{r}, \tag{1}$$

$$t(x) = T(x) - \frac{1}{2}\left[T^2(x) - T(x^2)\right], \tag{2}$$

$$\overline{H}(x) = \frac{x^2}{1+x} \exp \sum_{r=1}^{\infty} \frac{\overline{H}(x^r)}{r x^r}, \tag{3}$$

$$H(x) = \frac{1+x}{x}\overline{H}(x) - \frac{1}{2x}\left[\overline{H}^2(x) + \overline{H}(x^2)\right], \tag{4}$$

$$h(x) = H(x) - \frac{1}{2x^2}\left[\overline{H}^2(x) - \overline{H}(x^2)\right], \tag{5}$$

$$U(x) = x \exp \sum_{r=1}^{\infty} (-1)^{r+1} \frac{U(x^r)}{r}, \tag{6}$$

*e-mail: `marko.petkovsek@uni-lj.si`

†e-mail: `tomaz.pisanski@uni-lj.si`

$$i(x) \quad = \quad U(x) - \frac{1}{2}\left[U^2(x) + U(x^2)\right] . \tag{7}$$

These equations have been used to compute the numbers T_n, t_n, h_n, and i_n by expanding their left-hand and right-hand sides to order n about the point $x = 0$. Notice that this gives a recurrence for these sequences (or expresses them explicitly in terms of other known sequences).

Our function `FunEq` automates this process. Given an equation of the form `lhs == rhs` with unknown function `f[x]`, it sets up rules for computing derivatives of `f[x]` at $x = 0$. `FunEq` is based on the following assumptions:

1. `f[x]` is expressible in the form $f(x) = \sum_{n=0}^{\infty} c_n x^n$,
2. for all $n \geq 0$, the expansion of `lhs - rhs` to order n depends on c_n but not on any c_m with $m > n$,
3. for all $n \geq 0$, the expansion of `lhs - rhs` to order n is not affected by replacing infinite summation and/or infinite product bounds with n.

Although these requirements seem to be rather restrictive they are in fact satisfied (or satisfiable) in many cases of interest.

```
Mathematica 2.0 for MS-DOS 386/7 (June 21, 1991)
Copyright 1988-91 Wolfram Research, Inc.

In[1]:= SetOptions[$Output, PageWidth -> 55];

In[2]:= FunEq[lhs_ == rhs_, f_[x_]] :=
           Module[{c},
              f[0] := c[0];
              Derivative[n_][f][0] := n! c[n];
              c[n_] := c[n] = Block[{g, t, a},
                 g = Evaluate[Sum[c[k] #^k, {k,0,n-1}]
                      + a #^n] &;
                 t = Coefficient[Series[(lhs - rhs) /.
                      {Infinity -> n, f -> g}, {x,0,n}],
                      x, n];
                 a /. First[Solve[t == 0, a]] ]]
```

After executing `FunEq[lhs == rhs, f[x]]`, the Taylor coefficients of `f[x]` can be readily obtained by using `Derivative`, `Series`, or related functions.

```
In[3]:= FunEq[T[x] == x Exp[Sum[T[x^r]/r,
                                {r, Infinity}]], T[x]];
FunEq[t[x] == T[x] - (T[x]^2 - T[x^2])/2, t[x]];
FunEq[H1[x] == x/(1+x) Exp[Sum[H1[x^r]/r,
                               {r, Infinity}]], H1[x]];
FunEq[H[x] == (1+x) H1[x] - x (H1[x]^2 + H1[x^2])/2,
                                                 H[x]];
FunEq[h[x] == H[x] - (H1[x]^2 - H1[x^2])/2, h[x]];
FunEq[U[x] == x Exp[Sum[(-1)^(n+1)U[x^n]/n,
                                {n, Infinity}]], U[x]];
FunEq[i[x] == U[x] - (U[x]^2 + U[x^2])/2, i[x]]
```

For reasons of efficiency, we replaced $\overline{H}(x)$ by $H_1(x) = \overline{H}(x)/x$.

The table on page 232 of [1] gives T_n and t_n for $n \leq 26$, as well as h_n and i_n for $n \leq 12$. Also, T_n, t_n, h_n, and i_n are sequences no. 454, 299, 118, and 1022, respectively, of Sloane [3] where they are listed up to $n = 24$, 26, 32, and 12, respectively. We computed these sequences for $n \leq 30$. The timings were obtained on a 33 MHz 486 PC.

```
In[4]:= CoefficientList[T[x] + O[x]^31, x] // Timing

Out[4]= {141.11  Second,

>    {0, 1, 1, 2, 4, 9, 20, 48, 115, 286, 719, 1842,

>     4766, 12486, 32973, 87811, 235381, 634847,

>     1721159, 4688676, 12826228, 35221832, 97055181,

>     268282855, 743724984, 2067174645, 5759636510,

>     16083734329, 45007066269, 126186554308,

>     354426847597}}
```

```
In[5]:= CoefficientList[t[x] + O[x]^31, x] // Timing

Out[5]= {13.35 Second,

>    {0, 1, 1, 1, 2, 3, 6, 11, 23, 47, 106, 235, 551,

>     1301, 3159, 7741, 19320, 48629, 123867, 317955,

>     823065, 2144505, 5623756, 14828074, 39299897,

>     104636890, 279793450, 751065460, 2023443032,

>     5469566585, 14830871802}}
```

Note that it does not take long to compute t_n when the values for T_n are known.

```
In[6]:= CoefficientList[h[x] + O[x]^31, x] // Timing

Out[6]= {157.36 Second,

>    {0, 1, 1, 0, 1, 1, 2, 2, 4, 5, 10, 14, 26, 42,

>     78, 132, 249, 445, 842, 1561, 2988, 5671, 10981,

>     21209, 41472, 81181, 160176, 316749, 629933,

>     1256070, 2515169}}

In[7]:= CoefficientList[i[x] + O[x]^31, x] // Timing

Out[7]= {150.88 Second,

>    {0, 1, 0, 0, 0, 0, 0, 1, 1, 3, 6, 15, 29, 67,

>     139, 310, 667, 1480, 3244, 7241, 16104, 36192,
```

```
>     81435, 184452, 418870, 955860, 2187664, 5025990,

>     11580130, 26765230, 62027433}}
```

Incidentally, by first obtaining negative values for h_n we discovered an incorrect negative sign in equation (15.48) and a missing divisor of 2 in equation (15.49) of [1].

No doubt these times could be significantly reduced by considering each equation individually and expanding subexpressions only to the minimum necessary order. However, we satisfied ourselves with a simple, general-purpose function which can be used to find the first few terms of a sequence defined by equations similar to those in (1) – (7).

For another example consider the *generalized Catalan numbers* ${}_pc_n$ which count p-ary trees with n interior nodes. Their generating function $c_p(x)$ satisfies the equation $xc_p(x)^p = c_p(x) - 1$. For a proof of this fact and for further properties and combinatorial interpretations of these numbers, see [2]. Below we use `FunEq` to find the first few ${}_pc_n$ for $p = 2$ (the usual Catalan numbers) and $p = 5$.

```
In[7]:= FunEq[x c[2][x]^2 == c[2][x] - 1, c[2][x]];
FunEq[x c[5][x]^5 == c[5][x] - 1, c[5][x]]

In[8]:= CoefficientList[c[2][x] + O[x]^16, x]

Out[8]= {1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862,

>    16796, 58786, 208012, 742900, 2674440, 9694845}

In[9]:= CoefficientList[c[5][x] + O[x]^12, x]

Out[9]= {1, 1, 5, 35, 285, 2530, 23751, 231880,

>    2330445, 23950355, 250543370, 2658968130}
```

References

[1] F. Harary, *Graph Theory*, Addison-Wesley, Reading, Mass., 1968 (third printing, 1972).

[2] P. Hilton, J. Pedersen, Catalan numbers, their generalization, and their uses, *Math. Intelligencer* **13** (1991) 64 – 75.

[3] N.J.A. Sloane, *A Handbook of Integer Sequences*, Academic Press, New York and London, 1973.