Annals of Discrete Mathematics 33 (1987) 241 - 260

SOME RECENT RESULTS ON THE REGISTER FUNCTION OF A BINARY TREE

Helmut PRODINGER

Technische Universität Wien,
A-1040 Wien, Austria

In particular I had to make use of curves which are continuous but which are so crinkly that they can not properly said to have a direction. I have already pointed out in my discussion of the Brownian motion that these curves had been more or less the stepchildren of mathematics and had been regarded as rather unnatural museum pieces, derived by the mathematician from abstract considerations, and with no true representation in physics. Here I found myself establishing an essential physical theory in which such curves played an indispensable role.

Norbert Wiener, *I am a mathematician*

1. Introduction

We propose here to show that a large class of enumeration problems concerning trees can now be solved rather easily and automatically. This is in contrast to the situation several years ago when even the simplest problems in this class caused serious problems.

To be more explicit, we deal with binary trees; let $\mathfrak{B}$ be the family of binary trees, then we have the formal equation

$$\mathfrak{B} = \square + \overset{\circ}{\underset{\mathfrak{B}\quad\mathfrak{B}}{\bigwedge}}$$

expressing the fact that a binary tree is either empty or consists of a root together with a left and right subtree, each one being itself a binary tree.

We define the size $|t|$ of a tree t to be the number of internal nodes of t; t_n denotes the number of trees of size n. We find immediately from the formal equation for the generating function $B(z)$ of trees:

$$\begin{aligned} B(z) &:= \sum_{n\geq 0} t_n z^n = \sum_{t\in\mathfrak{B}} z^{|t|} = 1 + zB^2(z) \\ &= \frac{1-\sqrt{1-4z}}{2z} = \sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n} z^n. \end{aligned}$$

We deal with the *register function* reg(t) of a tree t. This function is defined inductively as follows:

$$\begin{cases} \operatorname{reg}(\square)=0 \\ \operatorname{reg}\left(\begin{smallmatrix} & \circ & \\ & \wedge & \\ t_1 & & t_2 \end{smallmatrix}\right)=\begin{cases} 1+\operatorname{reg}(t_1) & \text{if } \operatorname{reg}(t_1)=\operatorname{reg}(t_2) \\ \max\{\operatorname{reg}(t_1),\operatorname{reg}(t_2)\} & \text{otherwise}. \end{cases} \end{cases}$$

This function is of relevance in Computer Science; for this we refer to [1].

For completeness we now cite all papers dealing with the register function: [1–3, 5, 7–13].

The recursive definition can be most easily visualised by labelling the nodes of the tree in a bottom-up-sense; the value of the root is then the desired value reg(t). (Figure 1.)

The marked nodes cause the register function to increase; we call them *critical*. If we forget about all other nodes, we obtain the (*ordered binary*) *forest of critical nodes*, in the example presented in Figure 2.

In the next sections we deal with several enumeration problems concerning the register function. The interest is not so much in these parameters itself, but in the methodology that we are going to point out in a few seconds. However, there are still unsolved problems, for instance, what is the average number of components of the forest just mentioned?

We are interested in *average values* of certain parameters, where all trees of size n are to be considered equally likely. Since these values are of an intrinsic complexity, we confine ourselves to the determination of *asymptotic equivalents*; it will turn out that these contain *periodic fluctuations*, even if we do not compute them in all the examples. Using appropriate generating functions, the average value is

$$\frac{[z^n]E(z)}{[z^n]B(z)}$$

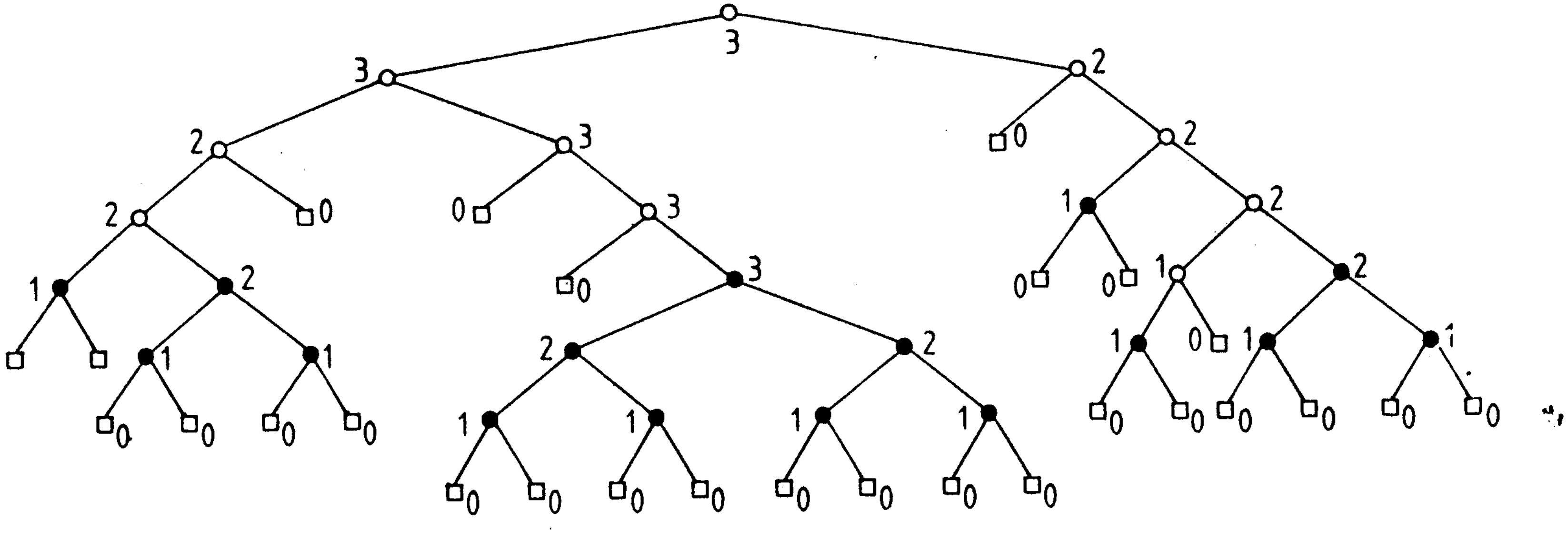

Figure 1.

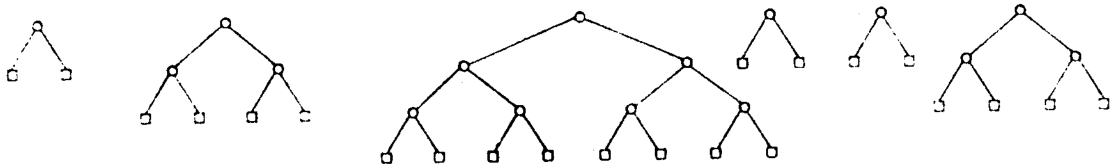

Figure 2.

where the notion $[z^n]f(z)$ refers to the coefficient of z^n in the Taylor expansion of the generating function $f(z)$. Such an $E(z)$ is often

$$E(z)=\sum_{p\geq 1} E_p(z) \quad \text{or} \quad E(z)=\sum_{p\geq 0} pE_p(z),$$

depending on what $E_p(z)$ counts.

So the first step is to find explicit expressions for the $E_p(z)$'s, whatever they might be, mostly by setting up recursions. Here, we use $R_p(z)$, the generating functions of trees with register function $=p$.

The next rule is to perform the change of variable $\boxed{z=u/(1+u)^2}$ in all generating functions. Here are some special functions:

$$B(z)=1+u, \quad \sqrt{1-4z}=\frac{1-u}{1+u}, \quad \frac{dz}{du}=\frac{1-u}{(1+u)^3},$$

$$R_p(z)=\frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}}, \quad S_p(z):=\sum_{j\geq p} R_j(z)=\frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^p}}.$$

Remark that $S_p(z)$ is the generating function of trees with register function $\geq p$. The explicit values for $R_p(z)$ and $S_p(z)$ appear for the first time in [3] and [8]; an alternative (and easy) proof can be found in [12]. In Section 7 a further proof is sketched.

Then it usually turns out that $E(z)$ is a linear combination of

$$f(u)\sum_{p\geq 1}\omega_p u^p$$

where $f(u)$ is a rational function and ω_p are defined by some arithmetic properties. Each natural number n can be uniquely written as $n=2^m(1+2j)$, $m,j\geq 0$; in a lot of cases ω_n is a function of m and j. The *corresponding Dirichlet series*

$$\sum_{p\geq 1}\omega_p p^{-s}$$

has then a closed form expression.

One is interested in a local expansion of $E(z)$ in a neighbourhood of its (unique) singularity at the radius of convergence at $z=1/4$. This expansion can be nicely written in terms of $\varepsilon(z)=\sqrt{1-4z}$ for $\varepsilon\to 0$.

In this paper all expansions can be carried out to any desired degree of accuracy; henceforth we never worry about O-terms. In other words, if we have computed that a certain coefficient in a power series expansion is asymptotic to a certain quantity and we want to get the order of growth of the difference, we simply get it by computing one more term in the asymptotic expansion.

$z \to 1/4$ means $u \to 1$, or with $u = e^{-t}$ it means $t \to 0$. The desired local expansion is obtained in terms of t. However, we can easily rewrite it in terms of ε, since

$$t \sim 2\varepsilon + \ldots .$$

The expansion of the rational function $f(u)$ is trivial.

For the sum one uses the *Mellin integral transform*. For details of this particularly useful feature we refer to a forthcoming book of Flajolet, Régnier and Sedgewick [6]

$$\mathfrak{M} V(t) := V^*(s) := \int_0^\infty t^{s-1} V(t)\, dt.$$

Since $V^*(at) = a^{-s} V^*(s)$, we see that

$$\mathfrak{M} \sum_p \omega_p e^{-tp} = \sum_p \omega_p p^{-s}\, \mathfrak{M} e^{-t},$$

so that the transform of the series in question is the product of the associated Dirichlet series and the classical gamma function $\Gamma(s)$. The Mellin inversion formula allows one to recover the original function:

$$V(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} V^*(s) t^{-s}\, ds,$$

where c is some appropriate real constant. If we shift the line of integration to the left, taking the residues into account, we thereby obtain an asymptotic series of the function $V(t)$ for $t \to 0$. Since the integrand contains as ingredients only things like $\Gamma(s)$, $\zeta(s)$, $(2^s - 1)^{-1}$, etc., the computation of the residues is particularly easy; one just needs some special values of the functions involved [14]. Observe that $(2^s - 1)^{-1}$ has poles at $\chi_k = (2k\pi i)/\log 2$, whence the periodic fluctuations already mentioned.

Once the local expansion is obtained, we use *translation lemmas* (see [4]); they allow to go from the local expansion of the generating function to the asymptotic behaviour of the coefficients. For this one has to have a repertoire of known expansion, like

$$[z^n] \frac{1}{\sqrt{1-4z}} \sim \frac{1}{\sqrt{\pi}} 4^n n^{-1/2}$$

etc. The last step is then to divide by

$$[z^n]\,B(z)\sim\frac{1}{\sqrt{\pi}}4^n n^{-3/2}.$$

2. The register pathlength

Let us recall the (ordinary) pathlength of a binary tree: one considers all paths between the root and an internal node and counts the number of internal nodes on such a path. Summing all those numbers over all paths yields the *pathlength* of the tree.

If we count on each path only the *number of critical nodes*, we have defined in this way the *register path length* rpl(t) of the tree t.

In this section we are concerned with the average register pathlength, considering each tree of size n to be equally likely.

We have the following recursion: ($|t|$ stands for the number of nodes in the tree t.)

$$\mathrm{rpl}(\square)=0$$

$$\mathrm{rpl}\Bigg(\underset{t_1\quad t_2}{\overset{\circ}{\bigwedge}}\Bigg)=\mathrm{rpl}(t_1)+\mathrm{rpl}(t_2)$$

$$+\left\{\left|\underset{t_1\quad t_2}{\overset{\circ}{\bigwedge}}\right|\quad\text{if the root is critical}\right\}.$$

We introduce the obvious generating function

$$P(z)=\sum_t \mathrm{rpl}(t)\,z^{|t|}$$

and obtain by a direct translation of the recursion for rpl the following equation for the corresponding generating functions:

$$P(z)=2zB(z)P(z)+\sum_{p\geq 0}\left([z^n]\,zR_p^2(z)\right)nz^n$$

or

$$P(z)=\frac{1}{\sqrt{1-4z}}\,z\frac{d}{dz}\sum_{p\geq 0}zR_p^2(z)$$

$$=\frac{u}{1-u^2}\frac{(1+u)^3}{1-u}\sum_{p\geq 1}\left[-\frac{1-u^2}{u^2}\frac{u^{2^p}}{(1-u^{2^p})^2}+\frac{(1-u)^2}{u^2}\frac{2^p u^{2^p}(1+u^{2^p})}{(1-u^{2^p})^3}\right]$$

$$=-\frac{(1+u)^3}{u(1-u)}A(u)+\frac{(1+u)^2}{u}B(u),$$

with

$$A(u)=\sum_{p\geq 1}\frac{u^{2^p}}{(1-u^{2^p})^2}$$

and

$$B(u)=\sum_{p\geq 1}\frac{2^p u^{2^p}(1+u^{2^p})}{(1-u^{2^p})^3}.$$

Now

$$A(u)=\sum_{p\geq 1}\frac{u^{2^p}}{(1-u^{2^p})^2}=-\frac{u}{(1-u)^2}+\sum_{p,\,\lambda\geq 0}\lambda u^{\lambda 2^p}$$

$$=-\frac{u}{(1-u)^2}+\sum_{n\geq 1}\psi(n)u^n,$$

with

$$\psi(n)=\sum_{n=\lambda 2^p}\lambda.$$

Now write $n=2^m(1+2j)$ in a unique way:

$$\psi(n)=\sum_{p=0}^{m}2^{m-p}(1+2j)=(1+2j)(2^{m+1}-1)$$

$$=2n-(1+2j),$$

so that

$$A(u)=\frac{u}{(1-u)^2}-F(u),$$

with

$$F(u)=\sum_{n\geq 1}(1+2j)u^n.$$

In a similar way,

$$B(u)=-\frac{u(1+u)}{(1-u)^3}+\sum_{p\geqslant 0}2^p\sum_{\lambda\geqslant 0}\lambda^2u^{\lambda 2^p}$$
$$=-\frac{u(1+u)}{(1-u)^3}+\sum_{n\geqslant 1}\theta(n)u^n,$$

with

$$\theta(n)=\sum_{n=\lambda 2^p}2^p\lambda^2=\sum_{p=0}^{m}[2^{m-p}(1+2j)]^2 2^p$$
$$=(1+2j)^2 2^m(2^{m+1}-1)=2n^2-n(1+2j).$$

Hence

$$B(u)=\frac{u(1+u)}{(1-u)^3}-D(u),$$

with

$$D(u)=\sum_{n\geqslant 1}2^m(1+2j)^2u^n.$$

This gives us

$$P(z)=\frac{(1+u)^3}{u(1-u)}F(u)-\frac{(1+u)^2}{u}D(u).$$

Now let

$$C_p(u)=\sum_{n\geqslant 1}(1+2j)\,n^p u^n,$$

so that $C_0(u)=F(u)$, $C_1(u)=D(u)$.

The corresponding Dirichlet series $\hat{C}_p(s)$, which is obtained by replacing u^k by k^{-s} is then

$$\hat{C}_p(s)=\sum_{m,\,j\geqslant 0}[2^m(1+2j)]^{-s}(1+2j)[2^m(1+2j)]^p$$
$$=\sum_{m\geqslant 0}2^{m(p-s)}\sum_{j\geqslant 0}(1+2j)^{p+1-s}$$

$$=2\frac{2^{s-(p+1)}-1}{2^{s-p}-1}\zeta(s-(p+1)).$$

Now we set, as proposed, $u=e^{-t}$ and find quite easily

$$\frac{(1+u)^3}{u(1-u)}\sim\frac{8}{t} \quad \text{and} \quad \frac{(1+u)^2}{u}\sim 4.$$

To find the local behaviour of $C_p(u)$ as t tends to zero, we Mellin transform it:

$$\begin{aligned}\mathfrak{M}C_p(e^{-t})&=\mathfrak{M}\sum_{n\geq 1}e^{-tn}(1+2j)n^p\\&=\Gamma(s)\hat{C}_p(s).\end{aligned}$$

The Mellin inversion formula gives for some $c>p+2$:

$$C_p(e^{-t})=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Gamma(s)2\frac{2^{s-(p+1)}-1}{2^{s-p}-1}\zeta(s-(p+1))t^{-s}ds.$$

Shifting the line of integration to the left and collecting residues we find

$$C_p(e^{-t})\sim\tfrac{2}{3}(p+1)!\,t^{-p-2};$$

if $p=0$ we have to add the term (originating from the double pole)

$$\frac{\zeta(-1)}{\log 2}\log t.$$

In particular we have

$$C_0(e^{-t})\sim\frac{2}{3}t^{-2}-\frac{1}{12}\frac{1}{\log 2}\log t,$$

$$C_1(e^{-t})\sim\tfrac{4}{3}t^{-3}.$$

Hence

$$P(z)\sim-\frac{2}{3}\frac{1}{\log 2}\frac{\log t}{t}, \qquad t\to 0$$

$$\sim -\frac{1}{6}\frac{1}{\log 2}\frac{\log(1-4z)}{\sqrt{1-4z}}, \qquad z\to 1/4.$$

The coefficient of z^n in $P(z)$ is therefore asymptotic to

$$\frac{1}{6\log 2}\frac{4^n \log n}{\sqrt{\pi n}}.$$

We have to divide this quantity by $t_n \sim 4^n \pi^{-1/2} n^{-3/2}$ to obtain the average register pathlength:

Theorem 1. *The average register pathlength where all binary trees of size n are considered to be equally likely, is asymptotic to*

$$\tfrac{1}{6} n \log_2 n \quad \textit{as} \quad n\to\infty.$$

The lower order terms involve periodic functions in $\log_4 n$ *which could be determined if desired.*

3. The leaves determining the register function

Let us consider the forest of critical nodes. We want to count the number of nodes on the highest level (this level is the register function). Intuitively speaking, if this number is small, deletion of just a few nodes would decrease the register function. It is somehow more natural to extend the definition of a critical node to the leaves. Counting in this way gives exactly twice the number described first.

Let $[z^n w^m]\, Q_p(z, w)$ count the number of trees of size n and register function p with m leaves "on the maximal level", furthermore

$$N_p(z) := \frac{\partial}{\partial w} Q_p(z, w)\Big|_{w=1}, \qquad N(z) := \sum_{p\geq 0} N_p(z).$$

We easily obtain the recursion

$$Q_p(z, w) = 2zQ_p(z, w)\sum_{j<p} R_j(z) + zQ_{p-1}^2(z, w), \qquad p\geq 1,$$

$$Q_0(z, w) = w.$$

Hence

$$N_p(z)=2zN_p(z)\big(B(z)-S_p(z)\big)+2zN_{p-1}(z)R_{p-1}(z),$$

$$N_0(z)=1,$$

or

$$\begin{aligned}
N_p(z)&=N_{p-1}(z)\frac{2zR_{p-1}(z)}{1-2z\big(B(z)-S_p(z)\big)}\\
&=N_{p-1}(z)\frac{2u^{2^{p-1}}}{1+u^{2^p}}\\
&=\prod_{j=1}^{p}\frac{2u^{2^{j-1}}}{1+u^{2^j}}\\
&=\frac{2^p u^{2^p-1}}{(1-u^{2^{p+1}})/(1-u^2)}\\
&=\frac{1-u^2}{u}\,\frac{2^p u^{2^p}}{1-2^{p+1}}\,.
\end{aligned}$$

$$\begin{aligned}
N(z)&=\frac{1-u^2}{u}\sum_{p\geq 0}\frac{2^p u^{2^p}}{1-u^{2^{p+1}}}\quad{}^{1)}\\
&=\frac{1-u^2}{u}\sum_{p\geq 0}\sum_{\lambda\geq 0}2^p u^{2^p(1+2\lambda)}\\
&=\frac{1-u^2}{u}\sum_{n\geq 1}\chi(n)u^n,
\end{aligned}$$

with

$$\chi(n)=2^m \text{ if } n=2^m(1+2j).$$

Let

$$C(u)=\sum_{n\geq 1}\chi(n)u^n;$$

[1] Another way of reasoning is as follows: The register function p is the size of the largest "complete" binary tree; there is only one such tree in the forest. This complete tree has 2^p leaves, hence $N(z)=\sum_{p\geq 0}2^pR_p(z)$.

the corresponding Dirichlet series $\hat{C}(s)$ is then

$$\hat{C}(s)=\sum_{m,j\geq 0}2^m[2^m(1+2j)]^{-s}$$
$$=\frac{1}{2}\frac{2^s-1}{2^{s-1}-1}\zeta(s).$$

Therefore, as before,

$$C(e^{-t})=\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty}\Gamma(s)\frac{1}{2}\frac{2^s-1}{2^{s-1}-1}\zeta(s)t^{-s}ds.$$

We have a double pole at $s=1$ and simple poles at $s=1+(2k\pi i)/\log 2$, $k\in \mathbb{Z}\setminus\{0\}$. Since

$$\zeta(s)\sim\frac{1}{s-1}+\gamma,\ \Gamma(s)\sim 1-\gamma(s-1)\quad\text{as } s\to 1$$

we find

$$C(e^{-t})\sim -\frac{1}{2\log 2}t^{-1}\log t+\frac{3}{4}t^{-1}$$
$$+\frac{1}{2\log 2}\sum_{k\neq 0}\Gamma(1+\chi_k)\zeta(1+\chi_k)t^{-1-\chi_k}.$$

Since

$$\frac{1-u^2}{u}\sim 2t$$

we find

$$N(z)\sim -\log_2 t+\frac{3}{2}+\frac{1}{\log 2}\sum_{k\neq 0}\Gamma(1+\chi_k)\zeta(1+\chi_k)t^{-\chi_k}$$
$$\sim -\log_2\varepsilon+\frac{1}{2}+\frac{1}{\log 2}\sum_{k\neq 0}\Gamma(1+\chi_k)\zeta(1+\chi_k)\varepsilon^{-\chi_k}.$$

For the coefficient of z^n in $N(z)$ we have the asymptotic equivalent

$$\frac{1}{2\log 2}\frac{4^n}{n}+\frac{1}{\log 2}\sum_{k\neq 0}\Gamma(1+\chi_k)\zeta(1+\chi_k)4^n n^{-1+\chi_k/2}/\Gamma\left(\frac{\chi_k}{2}\right).$$

By the duplication formula for the gamma function we have

$$\Gamma(1+\chi_k)/\Gamma\left(\frac{\chi_k}{2}\right)=\frac{1}{\pi}\chi_k\Gamma\left(\frac{1+\chi_k}{2}\right).$$

We finally divide by $t_n \sim 4^n \pi^{-1/2} n^{-3/2}$ and have:

Theorem 2. *The average number of leaves on the maximum (register) level is asymptotic to*

$$\sqrt{n}\, G(\log_4 n), \quad n \to \infty$$

with a periodic function $G(x)$ with period 1 *and the Fourier expansion*

$$G(x)=\sum_{k\in Z} g_k e^{2k\pi i x}$$

where

$$g_0=\frac{\sqrt{\pi}}{2\log 2}$$

and

$$g_k=\frac{1}{\sqrt{\pi}\log 2}\chi_k\Gamma\left(\frac{1+\chi_k}{2}\right)\zeta(1+\chi_k), \quad k\neq 0.$$

4. The average register value of a node

We now sum all register values obtained by the labelling procedure described in the introduction and divide by $2n+1$ which is the total number of nodes. The average of this quantity is then to be computed.

We consider a binary tree where each leaf is labelled (marked) and replace it by $\sum_{p\geq 0} p\mathfrak{R}_p$. In terms of generating functions this means

$$W(z)=\sum_{n\geq 0}(n+1)t_n z^n \sum_{p\geq 0} pR_p(z)$$

where the coefficient of z^n in $W(z)$ is the sum of all register values in all trees of size n.

$$W(z)=\frac{1}{\sqrt{1-4z}}\sum_{p\geq 0} pR_p(z).$$

Since $\sum_{p\geq 0} pR_p(z)$ appeared already in the computation of the average register function, and

$$\frac{1}{\sqrt{1-4z}}=\frac{1+u}{1-u}\sim\frac{2}{t},$$

we have (compare [5]) with a constant K

$$W(z)\sim\frac{2}{t}\left[t\log_2 t+Kt+\frac{1}{\log 2}\sum_{k\neq 0} 2\Gamma(\chi_k)\zeta(\chi_k)t^{-\chi_k}+2\right]$$
$$\sim 2\log_2\varepsilon+K'+\frac{4}{\log 2}\sum_{k\neq 0}\Gamma(\chi_k)\zeta(\chi_k)\varepsilon^{-\chi_k}+\frac{2}{\varepsilon}$$

and so

$$[z^n]\,W(z)\sim-\frac{1}{\log 2}\frac{4^n}{n}+\frac{4}{\log 2}\sum_{k\neq 0}\Gamma(\chi_k)\zeta(\chi_k)\frac{4^n n^{\chi_k/2-1}}{\Gamma\left(\frac{\chi_k}{2}\right)}+\frac{2\cdot 4^n}{\sqrt{\pi n}}.$$

We divide it by $(2n+1)t_n\sim 2\cdot 4^n\pi^{-1/2}n^{-1/2}$ and obtain

Theorem 3. *The average register value of a node in a tree of size n is asymptotic to*

$$1+\frac{1}{\sqrt{n}}H(\log_4 n)$$

where $H(x)$ is periodic with period 1 *and the Fourier expansion*

$$H(x)=\sum_{k\in Z} h_k e^{2k\pi i x}$$

with

$$h_0=\frac{-\sqrt{\pi}}{\log 2}$$

and

5. The average number of registers to evaluate r arithmetic expressions

Corresponding to the title we consider an ordered forest of r binary trees (r fixed) with altogether n internal nodes ("size n"). The register function extends trivially by taking the maximum of the register functions of the r trees.

To compute the average we consider the obvious generating functions (compare the introduction):

$$\begin{aligned} E(z) &= \sum_{p\geq 1} B^r(z)-(B(z)-S_p(z))^r \\ &= (1+u)^r \sum_{p\geq 1} 1-\left(\frac{1-u^{2^p-1}}{1-u^{2^p}}\right)^r \\ &= (1+u)^r \sum_{p\geq 1} 1-\left(1+\frac{(1-u^{-1})u^{2^p}}{1-u^{2^p}}\right)^r \\ &= -(1+u)^r \sum_{\lambda=1}^{r} \binom{r}{\lambda}(1-u^{-1})^\lambda \sum_{p\geq 1} \frac{u^{2^p\lambda}}{(1-u^{2^p})^\lambda}. \end{aligned}$$

Let us consider

$$E_\lambda(u) = \sum_{p\geq 1} \frac{u^{2^p\lambda}}{(1-u^{2^p})^\lambda};$$

$$\begin{aligned} \mathfrak{M}E_\lambda(e^{-t}) &= \mathfrak{M} \sum_{p\geq 1} \frac{e^{-t\lambda 2^p}}{(1-e^{-t2^p})^\lambda} \\ &= \sum_{p\geq 1} 2^{-ps}\,\mathfrak{M} \frac{e^{-t\lambda}}{(1-e^{-t})^\lambda} \\ &= \frac{1}{2^s-1}\,\mathfrak{M}\, e^{-t\lambda} \sum_{m\geq 0} \binom{\lambda+m-1}{\lambda-1} e^{-mt} \\ &= \frac{\Gamma(s)}{2^s-1} \sum_{m\geq 1} \binom{m-1}{\lambda-1} m^{-s} \\ &= \frac{\Gamma(s)}{2^s-1}\,\frac{1}{(\lambda-1)!} \sum_{m\geq 1} \left(m^{\lambda-1}-\binom{\lambda}{2}m^{\lambda-2}+\dots\right) m^{-s} \\ &= \frac{\Gamma(s)}{2^s-1}\,\frac{1}{(\lambda-1)!}\left[\zeta(s-\lambda+1)-\binom{\lambda}{2}\zeta(s-\lambda+2)+\dots\right]. \end{aligned}$$

Using the Mellin inversion formula we find for $\lambda \geqslant 2$

$$E_\lambda(e^{-t}) \sim \frac{1}{2^\lambda - 1} t^{-\lambda} - \frac{1}{2^{\lambda-1}-1} \frac{\lambda}{2} t^{-\lambda+1} + \dots .$$

For $\lambda = 1$ there is a double pole "in the second term"; this has also been computed previously:

$$E_1(e^{-t}) \sim \frac{1}{t} + \frac{1}{2}\log_2 t - \frac{1}{2}\log_2 2\pi + \frac{1}{4} + \frac{\gamma}{2\log 2}$$
$$+ \frac{1}{\log 2} \sum_{k \neq 0} \Gamma(\chi_k)\zeta(\chi_k) t^{-\chi_k} .$$

Also

$$-(1+u)^r(1-u^{-1})^\lambda \sim (-1)^{\lambda+1} t^\lambda 2^r \left(1-(r-\lambda)\frac{t}{2}\right).$$

Therefore

$$E(z) \sim \sum_{\lambda=1}^{r} \frac{1}{2^\lambda - 1}(-1)^{\lambda+1} 2^r \left(1-(r-\lambda)\frac{t}{2}\right)\binom{r}{\lambda}$$
$$- \sum_{\lambda=2}^{r} \frac{1}{2^{\lambda-1}-1} \frac{\lambda}{2} t 2^r (-1)^{\lambda+1} \binom{r}{\lambda}$$
$$+ rt2^r \left[\frac{1}{2}\log_2 t - \frac{1}{2}\log_2 2\pi + \frac{1}{4} + \frac{\gamma}{2\log 2}\right.$$
$$\left. + \frac{1}{\log 2}\sum_{k\neq 0} \Gamma(\chi_k)\zeta(\chi_k) t^{-\chi_k}\right].$$

The huge bracket is already well-known; from the two sums only the coefficient of t is of interest. It is

$$\sum_{\lambda=1}^{r} \frac{1}{2^\lambda - 1}(-1)^{\lambda+1} 2^r(-r+\lambda)\frac{1}{2}\binom{r}{\lambda}$$
$$- \sum_{\lambda=2}^{r} \frac{1}{2^{\lambda-1}-1} \frac{\lambda}{2} 2^r (-1)^{\lambda+1}\binom{r}{\lambda}$$
$$= 2^{r-1} \sum_{\lambda=1}^{r-1} \frac{(-1)^\lambda}{2^\lambda - 1}\left[(r-\lambda)\binom{r}{\lambda} - (\lambda+1)\binom{r}{\lambda+1}\right] = 0.$$

Since

$$B^r(z) \sim (2-2\varepsilon)^r \sim 2^r - r2^r\varepsilon,$$

we see that the desired average is (with respect to the first two terms in the asymptotic expansion!) independent of r. Hence we have

Theorem 4. *The average number of registers to evaluate r arithmetic expressions of (altogether) size n is asymptotic to*

$$\log_4 n + D(\log_4 n), \qquad n \to \infty$$

with the well-known function $D(x)$, being periodic with period 1.

6. How early is the register function reached

If we regard the labelling procedure which yields the register function at the root, it normally happens that this value appears already earlier. We now count the number of nodes above this first occurrence; this number is $\geqslant 0$. In this section we are concerned with the average of this parameter.

Let $Q_p(z, w)$ be the generating function of trees with register function p where the coefficient of w^j refers to the value j of our desired parameter. Furthermore let $N_p(z) = \frac{\partial}{\partial w} Q_p(z, w)\Big|_{w=1}$ and $N(z) = \sum_{p \geqslant 0} N_p(z)$. Naturally, we are interested in

$$\frac{[z^n]\,N(z)}{[z^n]\,B(z)}.$$

If we regard the original recursion for the register function reg and translate it into an equation of generating functions we obtain:

$$Q_p = 2zwQ_p \sum_{j<p} R_j + zR_{p-1}^2, \qquad p \geqslant 1,\ Q_0 = 1$$

and therefore

$$\begin{aligned} N_p &= 2z\,R_p(B-S_p) + 2z\,N_p(B-S_p) \\ &= \frac{2z\,R_p(B-S_p)}{1-2z\,(B-S_p)} \end{aligned}$$

$$=\frac{2(1+u)u^{2^p}(1-u^{2^{p-1}})}{(1-u^{2^p})(1+u^{2^p})^2}$$

$$=\frac{2(1+u)u^{2^p}}{(1-u^{2^{p+1}})^2}-\frac{2(1+u)^2}{u}\frac{u^{2^{p+1}}}{(1-u^{2^{p+1}})^2}+\frac{2(1+u)}{u}\frac{u^{3.2^p}}{(1-u^{2^{p+1}})^2}.$$

Let us consider

$$M_a(u)=\sum_{p\geqslant 0}\frac{u^{a.2^p}}{(1-u^{2^{p+1}})^2},$$

then

$$N(z)=2(1+u)M_1-\frac{2(1+u)^2}{u}M_2+\frac{2(1+u)}{u}M_3.$$

$$\mathfrak{M}\,M_a(e^{-t})=\mathfrak{M}\sum_{p\geqslant 1}\frac{e^{-t\frac{a}{2}2^p}}{(1-e^{-t2^p})^2}$$

$$=\sum_{p\geqslant 1}2^{-ps}\mathfrak{M}\frac{e^{-t\frac{a}{2}}}{(1-e^{-t})^2}$$

$$=\frac{1}{2^s-1}\mathfrak{M}\,e^{-t\frac{a}{2}}\sum_{m\geqslant 0}(m+1)e^{-mt}$$

$$=\frac{\Gamma(s)}{2^s-1}\sum_{m\geqslant 1}\left(m+\frac{a-2}{2}\right)^{-s}m.$$

Apart from the $\Gamma(s)/(2^s-1)$-factor we obtain for

$$a=1:\quad (2^{s-1}-1)\zeta(s-1)+\tfrac{1}{2}(2^s-1)\zeta(s)$$

$$a=2:\quad \zeta(s-1)$$

$$a=3:\quad (2^{s-1}-1)\zeta(s-1)-\tfrac{1}{2}(2^s-1)\zeta(s).$$

Applying the Mellin inversion formula again we find that $M_a(u)$ is asymptoti to

$$\sim\frac{1}{3}t^{-2}+\frac{1}{2}t^{-1}+\frac{\zeta(-1)}{2\log 2}\log t\qquad (a=1)$$