Philippe Flajolet and the Register function

Helmut Prodinger

Stellenbosch

December 15, 2011

Philippe Flajolet, J.-C. Raoult, and J. Vuillemin. The number of registers required to evaluate arithmetic expressions. Theoretical Computer Science, 9:99-125, 1979.

journal version; there is an earlier 1977 version

Philippe Flajolet, J.-C. Raoult, and J. Vuillemin. The number of registers required to evaluate arithmetic expressions. Theoretical Computer Science, 9:99-125, 1979.
journal version; there is an earlier 1977 version

Introductory remarks:

The master of the Mellin transform did not use it here!

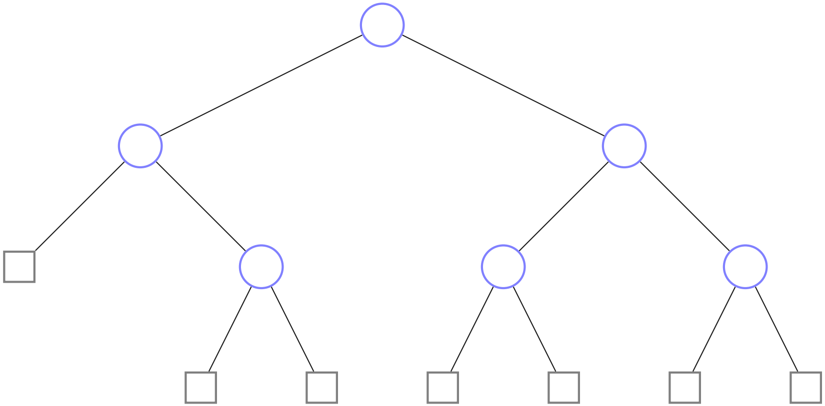
He often told the story that he learnt it from Rainer Kemp, who used Knuth's "Gamma-function method". (A special case of the method.)

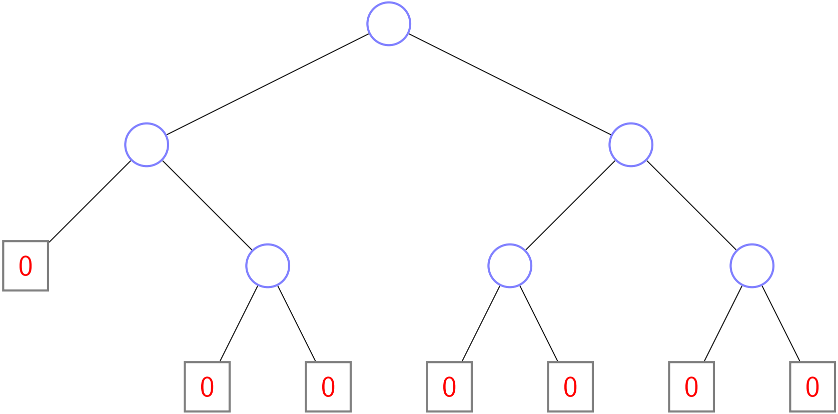
Here, everything is based on an elementary result about the summatory function of the sum-of-digits function, due to Delange.

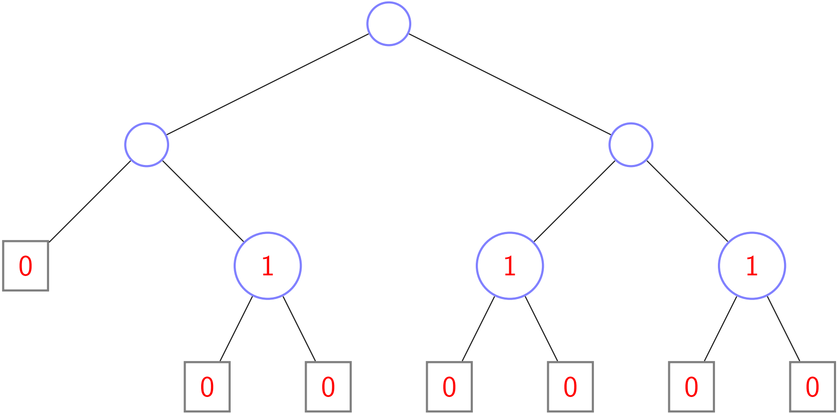
Introductory remarks:
The master of the Mellin transform did not use it here!
He often told the story that he learnt it from Rainer Kemp, who used Knuth's "Gamma-function method". (A special case of the method.)
Here, everything is based on an elementary result about the summatory function of the sum-of-digits function, due to Delange.

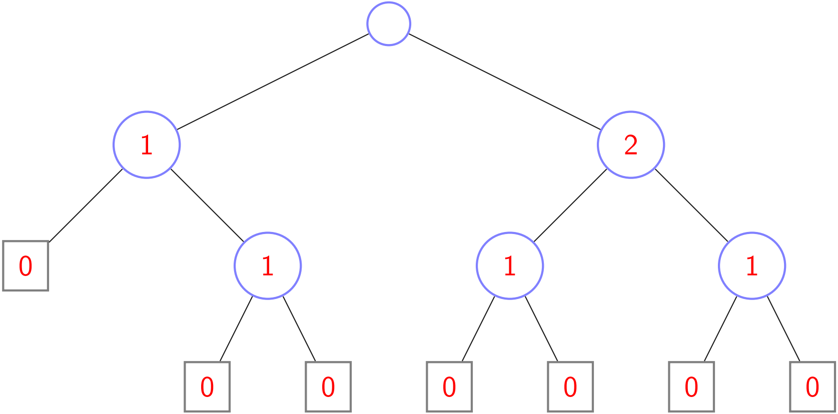
$\text{reg}(\square) = 0$, and if tree t has subtrees t_1 and t_2, then

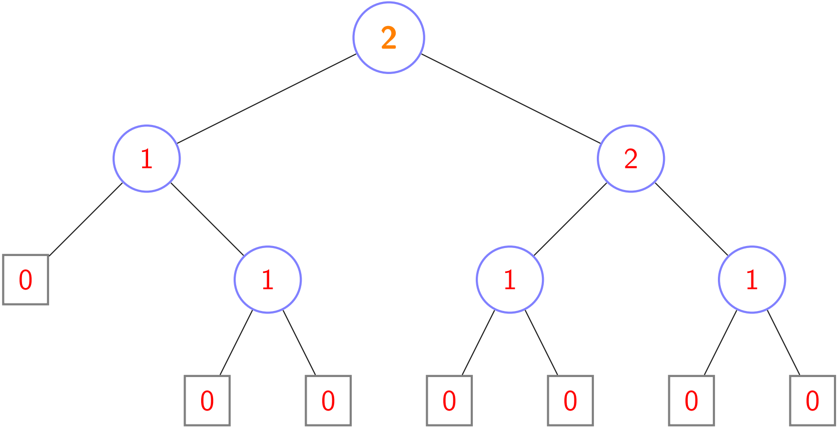
$$\begin{cases} \max\{\text{reg}(t_1), \text{reg}(t_2)\} & \text{if } \text{reg}(t_1) \neq \text{reg}(t_2), \\ 1 + \text{reg}(t_1) & \text{otherwise} \end{cases}$$




0
0
0
0
0
0
0
0


0
1
1
1
0
0
0
0
0
0


1
2
0
1
1
1
0
0
0
0
0
0


2
1
2
0
1
1
1
0
0
0
0
0
0

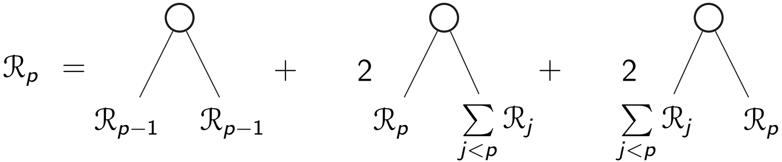
Flajolet liked symbolic equations!
Let $\mathcal{R}_p$ denote the family of trees with register function $= p$, then

In terms of generating functions:

$$R_p(z) = zR_{p-1}^2(z) + 2zR_p(z)\sum_{j<p} R_j(z)$$

Flajolet liked symbolic equations!

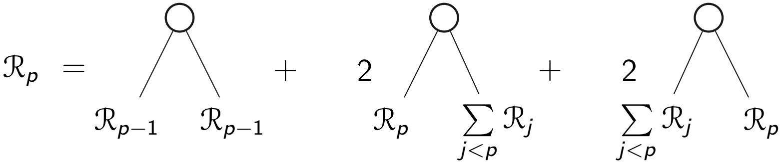
Let $\mathcal{R}_p$ denote the family of trees with register function $= p$, then



In terms of generating functions:

$$R_p(z) = zR_{p-1}^2(z) + 2zR_p(z)\sum_{j<p} R_j(z)$$

Flajolet liked symbolic equations!
Let $\mathcal{R}_p$ denote the family of trees with register function $= p$, then



In terms of generating functions:

$$R_p(z) = zR_{p-1}^2(z) + 2zR_p(z)\sum_{j<p} R_j(z)$$

Amazingly, this can be solved explicitly.
After some manipulations, a recursion pops up that is reminiscent of Chebyshev polynomials.

First, a trigonometric substitution was used, but eventually the one that De Bruijn, Knuth, and Rice also used:

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

Then

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

First, a trigonometric substitution was used, but eventually the one that De Bruijn, Knuth, and Rice also used:

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

Then

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

Reading off coefficients, the average number of registers requires to evaluate

$$\sum_{k\geq 1} v_2(k)\left[\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k}\right]$$

with $v_2(k)$ being the number of trailing zeroes in the binary representation of k.

The average number of registers to evaluate a binary tree with n nodes is asymptotically given by

$$\log_4 n + D(\log_4 n) + o(1)$$

with

$$D(x) = \sum_{k \in \mathbb{Z}} d_k e^{2\pi i k x}$$

and

$$d_0 = \frac{1}{2} - \frac{\gamma}{2\log 2} - \frac{1}{\log 2} + \log_2 \pi,$$
$$d_k = \frac{1}{\log 2} \zeta(\chi_k) \Gamma(\tfrac{\chi_k}{2})(\chi_k - 1),$$

with $\chi_k = \frac{2\pi i k}{\log 2}$. Perhaps Flajolet's first periodic oscillation?

a) The early Flajolet. (aka F-Raoult-Vuillemin)
Double summation:

$$\sum_{j \le k} v_2(j) = k - S_2(k)$$

$S_2(k)$ is the number of ones in the binary expansion of k.
It is known:

$$\sum_{m<n} S_2(m) = \frac{n \log_2 n}{2} + nF(\log_2 n).$$

This was shown by Delange and apparently mentioned to Flajolet directly.
The periodic function $F(t)$ is fully explicit in terms of Fourier coefficients.

A negative side effect of this is that the second difference of binomial coefficients became a fourth difference. No problem: approximations are available (Hermite polynomials).

$$\sum_{k\geq 1}\left[\frac{k\log_2 k}{2}+kF(\log_2 k)\right]H_4\Big(\frac{k}{\sqrt{n}}\Big)e^{-k^2/n}$$

$$\sum_{k\geq 1}\left[\frac{k\log_2 k}{2} + kF(\log_2 k)\right]H_4\left(\frac{k}{\sqrt{n}}\right)e^{-k^2/n}$$

This is doable (Riemann sums, controlling the error) but a bit dry.
But: completely elementary!

The mergesort recurrence

$$f(n) = f(\lfloor \tfrac{n}{2} \rfloor) + f(\lceil \tfrac{n}{2} \rceil) + \lfloor \tfrac{n}{2} \rfloor$$

is solved by

$$f(n) = \sum_{m<n} S_2(m).$$

Since Flajolet developed a calculus how to solve such recursions (Golin's talk), he got as a bonus a quick derivation of Delange's result.

Remark. To solve explicitly for R_p is somewhat crucial. Duchon et al. suggested a generalisation where no explicit formula is available, and Drmota and myself could only identify the leading $\log_4 n$ term! Delange's paper was extended and generalized into many different directions by many people.

Remark. To solve explicitly for R_p is somewhat crucial. Duchon et al. suggested a generalisation where no explicit formula is available, and Drmota and myself could only identify the leading $\log_4 n$ term! Delange's paper was extended and generalized into many different directions by many people.

b)
After Flajolet learnt about the Mellin transform, he attacked a sum like

$$\sum_{k\geq 1} v_2(k) H_2(kt) e^{-k^2 t^2}$$

$(t = 1/\sqrt{n})$ directly.
This goes well, since $v_2(2k+1) = 0$ and $v_2(2k) = 1 + v_2(k)$ and

$$\sum_{k\geq 1} \frac{v_2(k)}{k^s} = \frac{\zeta(s)}{2^s - 1}.$$

This appears already in his *thèse d'etat*.

c) After Flajolet became familiar with *singularity analysis of generating functions*, thanks to A. Odlyzko, he would consider

$$E(z) = \sum_{p \geq 1} pR_p(z) = \sum_{p \geq 1} p \frac{1 - u^2}{u} \frac{u^{2^p}}{1 - u^{2^{p+1}}}$$

and study it around the singularity $u = 1$ with ...

The Mellin transform!

But now on the level of the generating function itself, not the coefficients.

c) After Flajolet became familiar with *singularity analysis of generating functions*, thanks to A. Odlyzko, he would consider

$$E(z) = \sum_{p \geq 1} p R_p(z) = \sum_{p \geq 1} p \frac{1-u^2}{u} \frac{u^{2^p}}{1-u^{2^{p+1}}}$$

and study it around the singularity $u = 1$ with ...
The Mellin transform!
But now on the level of the generating function itself, not the coefficients.

c) After Flajolet became familiar with *singularity analysis of generating functions*, thanks to A. Odlyzko, he would consider

$$E(z) = \sum_{p \geq 1} p R_p(z) = \sum_{p \geq 1} p \frac{1 - u^2}{u} \frac{u^{2^p}}{1 - u^{2^{p+1}}}$$

and study it around the singularity $u = 1$ with ...
The Mellin transform!
But now on the level of the generating function itself, not the coefficients.

I studied last week his paper with Bruce Richmond, and he uses the strategy "Mellin, followed by singularity analysis" also in the context of b-digital search trees.

$$\sum_{p\geq 1} p \frac{u^{2^p}}{1-u^{2^{p+1}}}$$

or $u = e^{-t}$ $(u \sim 1 \leftrightarrow t \sim 0)$

$$\sum_{p\geq 1} p \frac{e^{-t2^p}}{1-e^{-t2^{p+1}}}$$

$$\sum_{p\geq 1,\ \lambda\geq 0} p e^{-t2^p(1+2\lambda)} = \sum_{n\geq 1} v_2(n) e^{-tn}$$

This is a harmonic sum!

A local expansion around $t \sim 0$ is thus found. It translates:

$$z \sim \frac{1}{4} - \frac{1}{16} t^2 + \dots$$

or

$$\sqrt{1 - 4z} \sim 2t$$

$$K = -1 + \log_2 \pi + \frac{1}{2} + \frac{\gamma}{\log 2}$$

$$r = \sqrt{1 - 4z}$$

$$E(z) = 2r \log_2 r + 2(K+1)r + 4 \sum_{k \neq 0} c_k r^{1-\chi_k} + \dots$$

$$c_k = \frac{1}{\log 2} \zeta(\chi_k) \Gamma(\chi_k).$$

This can be translated into an asymptotic expansion of the coefficients.

$$K = -1 + \log_2 \pi + \frac{1}{2} + \frac{\gamma}{\log 2}$$

$$r = \sqrt{1 - 4z}$$

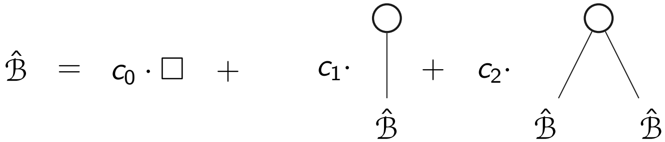
$$E(z) = 2r \log_2 r + 2(K+1)r + 4 \sum_{k \neq 0} c_k r^{1-\chi_k} + \dots$$

$$c_k = \frac{1}{\log 2} \zeta(\chi_k) \Gamma(\chi_k).$$

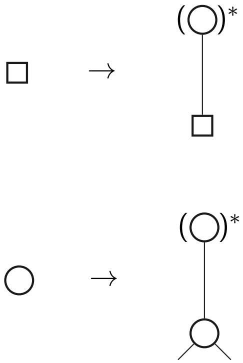
This can be translated into an asymptotic expansion of the coefficients.

With very little extra effort this can be used to treat unary-binary trees:
(paper by Flajolet/Prodinger)
Unary node: register function does not increase.

With very little extra effort this can be used to treat unary-binary trees:
(paper by Flajolet/Prodinger)
Unary node: register function does not increase.



$\hat{\mathcal{B}} = c_0 \cdot \square + c_1 \cdot$
$+ c_2 \cdot$
$\hat{\mathcal{B}}$
$\hat{\mathcal{B}}$
$\hat{\mathcal{B}}$



$$yB(yz)$$

marks internal nodes and leaves.

$$y \to \frac{c_0 y}{1 - c_1 z}$$

$$z \to \frac{c_2 z}{1 - c_1 z}$$

Options for size:

Count both, leaves and internal nodes

Count only internal nodes

Options for size:
Count both, leaves and internal nodes
Count only internal nodes

Options for size:
Count both, leaves and internal nodes
Count only internal nodes

Not much changes in the result:

$$\log_4 n + D(\log_4 n - \text{constant}) - \text{constant}$$

A NOTE ON GRAY CODE AND ODD-EVEN MERGE
P. FLAJOLET AND LYLE RAMSHAW
"Note" has 17 pages!
Gray code: Pattern last digit: 0110 0110 0110 ...
penultimate: 00111100 00111100 00111100 ...
and so on.
Formula:

$$a_k(n) = \left\lfloor \frac{n}{2^{k+2}} + \frac{3}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac{1}{4} \right\rfloor$$

A NOTE ON GRAY CODE AND ODD-EVEN MERGE
P. FLAJOLET AND LYLE RAMSHAW
"Note" has 17 pages!
Gray code: Pattern last digit: 0110 0110 0110 ...
penultimate: 00111100 00111100 00111100 ...
and so on.
Formula:

$$a_k(n) = \left\lfloor \frac{n}{2^{k+2}} + \frac{3}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac{1}{4} \right\rfloor$$

Delange type approach works, since

$$\left\lfloor \frac{n}{2^{k+2}} + \frac{3}{4} \right\rfloor = \int_n^{n+1} \left\lfloor \frac{t}{2^{k+2}} + \frac{3}{4} \right\rfloor dt$$

The quantity of interest (Sedgewick, odd-even merge):

$$1+(n+1)\sum_{i\geq 1}\beta(i)\frac{\binom{2n}{n+i+2}-3\binom{2n}{n+i+1}+3\binom{2n}{n+i}-\binom{2n}{n+i-1}}{\binom{2n}{n}},$$

with $\beta(i)$ being the number of ones in the GRAY code representation of i.

All 3 approches (elementary, Mellin, Mellin+singularity analysis) are available in this instance as well.

The quantity of interest (Sedgewick, odd-even merge):

$$1+(n+1)\sum_{i\geq 1}\beta(i)\frac{\binom{2n}{n+i+2}-3\binom{2n}{n+i+1}+3\binom{2n}{n+i}-\binom{2n}{n+i-1}}{\binom{2n}{n}},$$

with $\beta(i)$ being the number of ones in the GRAY code representation of i.
All 3 approches (elementary, Mellin, Mellin+singularity analysis) are available in this instance as well.

In a paper with coauthors Grabner, Kirschenhofer, Prodinger, Tichy, he used the Mellin-Perron technique to deal with digital sums.
More exotic things could also be handled with this approach: Stein's suggestion to interpret a binary expansion as a ternary expansion, representations of integers of sums of 3 squares, etc.